V. CRAIU

VERIFICAREA IPOTEZELOR STATISTICE

EDITURA DIDACTICĂ ȘI PEDAGOGICĂ
BUCUREȘTI, 1972

MINISTERUL EDUCAȚIEI ȘI ÎNVĂȚĂMÎNTULUI

VIRGIL CRAIU

VERIFICAREA IPOTEZELOR STATISTICE

EDITURA DIDACTICĂ ȘI PEDAGOGICĂ
BUCUREȘTI

Scanned with OKEN Scanner

PREFAȚĂ

Cu fiecare zi viața este influențată, din ce în ce mai mult, de deciziile bazate pe informații cantitative. Ipotezele științifice, experiențele și verificarea ipotezelor sînt acum familiare în studiul problemelor din fiecare domeniu de activitate. Totodată, metodele statistice moderne, bazate pe teoria probabilităților, sînt de neînlocuit în fizică și biologie, în economie și în sociologie, în medicină și agricultură, în psihologie și educație, în industrie și în administrația de stat. Astronomul stabilește următoarea poziție a unei comete pe baza metodelor statistice; asigurările își bazează activitatea pe tabelele de mortalitate și pe înregistrările statistice; cercetătorii din agricultură și medicină stabilesc dacă o nouă metodă, un nou tratament este superior pe baza considerațiilor statistice; dezvoltarea sistemului energetic nu poate avea loc fără cunoașterea datelor statistice asupra cererii de energie electrică; inginerii găsesc de neînlocuit teoria selecției în controlul calității producției: administrația de stat își bazează acțiunile pe datele statistice.

Cu toată această varietate de domenii în care se aplică, marea majoritate a metodelor statistice sînt aceleași, diferit fiind numai modul de aplicare.

Se înțelege că cercetătorul (dintr-un anumit domeniu de activitate) nu-și va sprijini decît foarte rar, concluziile numai pe o metodă (un test) de verificare a unei anumite ipoteze. Cunoscînd că o probă furnizează argumente, dar nu dovezi, el va lua în considerare alături de argumentele deja acumulate prin experiență și pe acesta. De asemenea există o informație colaterală care se amplifică în cursul experienței. În sfîrșit, dacă experiența prezintă o importanță mai mare, ea duce la noi ipoteze, care trebuie verificate cu ajutorul altor experiențe sau prin concordanța lor cu structura cunoscută și științei.

Scanned with OKEN Scanner

Cercetătorului îi revine responsabilitatea să adune aceste argumente și să ia o decizie. El nu poate scăpa de această responsabilitate prin citarea valorii unei anumite statistici (un X^2, un t, un F etc.).

Prin urmare, nici o metodă statistică prin ea însăși nu poate să împiedice greșeala, inexactitatea, raționamentul fals sau concluziile greșite. Datele originale trebuie să fie exacte, metodele trebuie adecvat aplicate și rezultatele trebuie interpretate de cineva care cunoaște bine nu numai metodele ci și domeniul în care ele sînt aplicate.

Metodele statistice pe care le vom prezenta pot fi considerate ca unelte care, în mîinile unor oameni pricepuți și aplicate la situații pentru care au fost create, dau rezultate folositoare.

Plecînd de la lema lui Neyman-Pearson, în capitolul 1, se construiesc regiuni critice pentru diferite ipoteze privind parametrii repartițiilor normale și se determină puterile corespunzătoare acestor teste.

În capitolul 2, folosind aproximația normală și transformările unghiulare, se verifică ipoteze privind parametrii repartițiilor binomiale. Se prezintă testul lui Fisher pentru compararea a două frecvențe observate. În continuare se prezintă teste pentru verificarea ipotezelor privind parametrii repartițiilor hipergeometrice și Poisson.

În capitolul 3 se prezintă testul X^2 și se dă un exemplu de folosire a acestuia la verificarea normalității.

În capitolul 4 se prezintă teste pentru egalitatea mediilor și dispersiilor. Se descriu testele: Bartlett, Cochran, t, F. De asemenea se prezintă testul lui Dixon privind eliminarea datelor ce diferă semnificativ de celelalte date. În încheiere se prezintă forme analoage testului t-necentrat. Se dau aplicații privind fabricarea anvelopelor și obținerea cuprului.

În capitolul 5 se prezintă cîteva dintre testele neparametrice și anume testul semnelor, testul medianei, testul iterațiilor, testul Wilcoxon, testul lui Kruskal și Wallis, testul X, testul lui Kolmogorov, testul lui Masey și testul lui Sarkady. Se dau aplicații privind durabilitatea sculelor (testul X, testul lui Masey), tenacitatea firelor de relon (testul lui Sarkady).

În capitolul 6 se face o prezentare succintă a analizei dispersionale uni și bifactoriale, capitolul conținînd și o aplicație privind planificarea experiențelor.

În capitolul 7 se determină, prin metoda celor mai mici pătrate, estimații pentru parametrii regresiei liniare, se compară două drepte de regresie, verificîndu-se ipoteze privind paralelismul dreptelor, iden-

Scanned with OKEN Scanner

titatea dreptelor; se prezintă regresia liniară cu ajutorul analizei dispersionale.

În capitolul 8 se estimează parametrii repartiției normale bidimensionale, se verifică ipoteza privind semnificația coeficientului de corelație, se dă un exemplu privind studiul deformării elastice, o metodă de reducere a numărului de probe în controlul statistic de recepție.

În capitolul 9 se studiază regresia lui y *asupra a* r *variabile independente* x_1.

În capitolul 10 se prezintă cîteva aspecte ale verificării ipotezelor cu ajutorul metodei secvențiale. Capitolul se încheie cu aplicații din controlul calității produselor.

La sfîrșitul cărții s-au plasat tabele strict necesare în abordarea problemelor practice.

Pentru înțelegerea în bune condiții a conținutului acestei cărți recomandăm cititorului să studieze în prealabil cărțile [4], [23] *și* [24].

Mulțumesc cercetătorilor D. V. Iliescu și V. Gh. Vodă (de la sectorul de Statistică Industrială din cadrul Centrului de Statistică Matematică al Academiei R.S.România) pentru sprijinul dat în definitivarea lucrării.

AUTORUL

Scanned with OKEN Scanner

TABLA DE MATERII

Cap. 1. Testul raportului de verosimilitate şi verificarea ipotezelor statistice privind parametrii repartiţiei normale. 9

Cap. 2. Verificări de ipoteze privind parametrii repartiţiilor binomială, hipergeometrică şi Poisson 26

Cap. 3. Testul X^2 51

Cap. 4. Teste pentru egalitatea dispersiilor şi mediilor 60

Cap. 5. Teste neparametrice 98

Cap. 6. Analiză dispersională 124

Cap. 7. Regresie lineară 139

Cap. 8. Repartiţia normală bidimensională şi coeficientul de corelaţie 161

Cap. 9. Regresie multiplă şi analiză dispersională 178

Cap. 10. Testul secvenţial 191

Bibliografie 203

Anexe 206

Tabelul 1. Valorile funcţiei $\Phi(l) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{l} e^{-\frac{x^2}{2}} dx$ 206

Tabelul 2. Valorile funcţiei inverse $\psi(x)$ a lui $\Phi(l)$ 208

Tabelul 3. Valorile lui χ^2 (Testul χ^2) 212

Tabelul 4. Valorile funcţiei $K(\lambda) = \sum_{-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$ 214

Tabelul 5,a. Valorile lui $F_{0,95}$ (testul F) 216

Tabelul 5,b. Valorile lui $F_{0,99}$ 218

Tabelul 5,c. Valorile lui $F_{0,975}$ (testul F) 220

Tabelul 6. Valorile lui t (testul Student) 221

Tabelul 7. Valorile lui Z, $Z = \frac{1}{2} \lg \frac{1+r}{1-r}$ 222

Tabelul 8. Valorile lui X_ε (testul lui van der Waerden) 223

Tabelul 9,a. Testul lui Cochran 225

Tabelul 9,b. Testul lui Cochran 226

Scanned with OKEN Scanner

CAPITOLUL

1

TESTUL RAPORTULUI DE VEROSIMILITATE ŞI VERIFICAREA IPOTEZELOR STATISTICE PRIVIND PARAMETRII REPARTIŢIEI NORMALE

§ 1. Noţiuni introductive. Ipoteze statistice. Regiune critică

Ştiinţele experimentale fac apel la metodele statisticii matematice îndeosebi privind verificarea unor ipoteze. De exemplu, în agricultură se introduce o nouă metodă privind cultura unei cereale oarecare, în tehnică o nouă metodă tehnologică de la care se aşteaptă o creştere a productivităţii şi îmbunătăţirii calităţii; în medicină un nou medicament etc. Se pune întrebarea, cum verificăm superioritatea metodei noi? Este noua metodă, noul produs superior celui anterior sau nu?

Cum să experimentăm noile procedee încît să ajungem la concluzia superiorităţii lor? Tot timpul ne putem pune întrebarea dacă rezultatele obţinute prin metodele noi utilizate nu se datoresc unor factori întîmplători. Acceptarea sau neacceptarea superiorităţii unei noi metode, dacă nu este studiată cu toată atenţia, ne poate conduce la concluzii neştiinţifice.

Pentru a veni în sprijinul ştiinţelor experimentale, statistica matematică a elaborat metode şi modele care ne permit să verificăm anumite ipoteze făcute.

O *ipoteză statistică* este o presupunere asupra uneia sau mai multor repartiţii ce caracterizează anumite populaţii şi mai precis asupra uneia sau mai multor parametri ai unor astfel de repartiţii. Ipoteza statistică este întotdeauna o presupunere asupra populaţiei şi nu asupra selecţiei.

Presupunerea este numită ipoteză deoarece se referă la o situaţie care poate fi adevărată.

Ipotezele statistice nu sînt aproape niciodată echivalente cu ipotezele ştiinţifice, care în mod obişnuit sînt presupuneri asupra fenomenelor.

În mod obişnuit ipoteza statistică este o descriere concretă a unora sau mai multor aspecte a uneia sau mai multor populaţii.

Scanned with OKEN Scanner

Să considerăm o repartiție unidimensională a cărei densitate de repartiție $p(x, \theta)$ depinde de un parametru θ și să verificăm ipoteza conform căreia parametrul θ are valoarea θ_0. Notăm această ipoteză astfel $H_0 : \theta = \theta_0$.

Să presupunem că în afară de valoarea θ_0, parametrul mai poate avea și una din valorile $\theta_1, \theta_2, \ldots$

Ipotezele

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1, \quad H_2 : \theta = \theta_2, \ldots$$

sînt numite *ipoteze admisibile.*

Pentru a o distinge de celelalte ipoteze, ipoteza H_0 este numită *ipoteza nulă.* Orice altă ipoteză admisibilă este numită *ipoteză alternativă.*

Fie $p(x, \theta_1, \theta_2, \ldots, \theta_k)$ densitatea unei repartiții ce depinde de parametrii necunoscuți $\theta_1, \theta_2, \ldots, \theta_k$.

Ipoteza $H: \theta_1 = \theta_{10}, \theta_2 = \theta_{20}, \ldots, \theta_k = \theta_{k0}$ care specifică valorile tuturor parametrilor necunoscuți este numită *ipoteză simplă.*

Fie $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ și Ω spațiul k-dimensional al tuturor vectorilor θ. Spațiul Ω este numit *spațiul parametrilor.*

Ipoteza $H: \theta \in \omega$, unde ω este o submulțime a lui Ω care conține cel puțin un punct, este numită *ipoteză compusă.*

Vom considera mai întîi cazul a două ipoteze admisibile. Fie $H_0 : \theta = \theta_0$ ipoteza nulă și $H_1 : \theta = \theta_1$ ipoteza alternativă.

Verificăm ipoteza H_0 cu ajutorul valorii observate a unei statistici u^*. Fie V mulțimea valorilor statisticii u astfel încît dacă ipoteza nulă este adevărată $P(u \in V) = \alpha$. Mulțimii V îi corespunde în spațiul de selecție o mulțime W astfel încît

$$P(\mathbf{x} \in W \mid H_0) = \alpha, \tag{1.1}$$

unde $\mathbf{x} = (x_1, \ldots, x_n)$ reprezintă vectorul de observații, iar valoarea numerică a lui α poartă numele de *pragul de semnificație al testului.*

Mulțimea W este numită *regiune critică.*

Dacă $\mathbf{x} \in W$ respingem ipoteza H_0 — echivalent aici cu acceptarea ipotezei H_1, iar dacă $\mathbf{x} \in \overline{W}$ ($\overline{W}$ complementara mulțimii W) acceptăm ipoteza H_0.

Mulțimea $\overline{W}$ este numită *regiune de acceptare.*

În general, putem găsi mai multe și adeseori chiar o infinitate de submulțimi, W ale *spațiului de selecție* (o mulțime de

---

* O funcție de observații.

Scanned with OKEN Scanner

elemente astfel încît orice eveniment corespunde unui singur element al acestei mulțimi), toate satisfăcînd (1.1).

Se pune problema alegerii unei anumite mulțimi W. Rezolvarea acestei probleme este dată în cele ce urmează.

§ 2. Funcția de putere. Cel mai puternic test

Există posibilitatea de a comite o eroare care constă în respingerea ipotezei H_0 cînd ea este adevărată. O astfel de eroare este numită *eroare de ordinul întîi.* Probabilitatea acestei erori este α, deci probabilitatea erorii de ordinul întîi este egală cu pragul de semnificație al testului. Natural dorim să facem o astfel de eroare cît mai rar. În general α se ia 0,01 sau 0,05.

Putem face și următoarea eroare: să admitem ipoteza H_0 cînd ea este falsă. O astfel de eroare este numită *eroare de ordinul al doilea.* Facem această eroare cînd $\mathbf{x} \notin W$, deși $\theta = \theta_1$. Notînd probabilitatea acestui eveniment prin β, avem

$$P(\mathbf{x} \in \overline{W} \mid \theta_1) = \beta \tag{2.1}$$

sau

$$P(\mathbf{x} \in W \mid \theta_1) = 1 - \beta. \tag{2.1'}$$

Probabilitatea de respingere a ipotezei H_0, ca funcție de θ poartă numele de *funcție de putere a testului.* O notăm prin $\pi(W, \theta)$. Deci

$$\pi(W, \theta) = P(\mathbf{x} \in W \mid \theta). \tag{2.2}$$

Din definiția lui α și β și din (2.2) obținem

$$\pi(W, \theta_0) = \alpha, \quad \pi(W, \theta_1) = 1 - \beta.$$

În controlul statistic al calității produselor, α poartă numele de *riscul furnizorului*, iar β *riscul beneficiarului.*

În cele ce urmează dăm regula de alegere a mulțimii W care servește ca bază a testului. Dintre toate mulțimile W pentru care (1.1) are loc, alegem ca bază a testului o mulțime W (dacă există) pentru care (2.1) are valoarea minimă.

O astfel de mulțime este numită *cea mai bună regiune critică*, iar testul bazat pe cea mai bună regiune critică este numit *cel mai puternic test*; nu există întotdeauna, însă poate fi construit pentru o clasă largă de funcții de repartiție.

Scanned with OKEN Scanner

Lema Neyman-Pearson. Fie $\mathbf{x} = (x_1, \ldots, x_n)$ o variabilă aleatoare de tip continuu (sau discret) a cărei densitate $P(\mathbf{x}, \theta)$ depinde de θ. Fie $P(\mathbf{x}, \theta_0)$ $[P(\mathbf{x}, \theta_1)]$ densitatea punctului aleator $\mathbf{x}$ dacă este adevărată ipoteza $H_0 : \theta = \theta_0 [H_1 : \theta = \theta_1]$. Fie $\mathfrak{R}$ mulţimea tuturor submulţimilor W ale spaţiului de selecţie pentru care

$$\sum_{\mathbf{x} \in W} P(\mathbf{x}, \theta_0) = \alpha \text{ (în cazul discret)} \tag{2.3}$$

sau

$$\int_W P(\mathbf{x}, \theta_0)\, d\mathbf{x} = \alpha \text{ (în cazul continuu)}, \tag{2.4}$$

unde $0 < \alpha < 1$. Presupunem că există $W^* \subset \mathfrak{R}$ astfel încît pentru $k \geqslant 0$,

$$\frac{P(\mathbf{x}, \theta_1)}{P(\mathbf{x}, \theta_0)} \geqslant k \tag{2.5}$$

pentru toate punctele $x \in W^*$ şi

$$\frac{P(\mathbf{x}, \theta_1)}{P(\mathbf{x}, \theta_0)} \leqslant k \tag{2.6}$$

pentru toate punctele $\mathbf{x} \notin W^*$. În aceste condiţii

$$\sum_{\mathbf{x} \in W^*} P(\mathbf{x}, \theta_1) \geqslant \sum_{\mathbf{x} \in W} P(\mathbf{x}, \theta_1) \tag{2.7}$$

sau

$$\int_{W^*} P(\mathbf{x}, \theta_1)\, d\mathbf{x} \geqslant \int_W P(\mathbf{x}, \theta_1)\, d\mathbf{x}, \tag{2.8}$$

pentru toţi $W \subset \mathfrak{R}$, adică W^* este cea mai bună regiune critică.

Vom da demonstrarea acestei leme în cazul discret. Fie $W \in \mathfrak{R}$ şi

$$\delta = \sum_{\mathbf{x} \in W^*} P(\mathbf{x}, \theta_1) - \sum_{\mathbf{x} \in W} P(\mathbf{x}, \theta_1)$$

Arătăm că $\delta \geqslant 0$. Deoarece

$$\sum_{\mathbf{x} \in W^*} P(\mathbf{x}, \theta_1) = \sum_{\mathbf{x} \in W^* \cap W} P(\mathbf{x}, \theta_1) + \sum_{\mathbf{x} \in W^* \cap \overline{W}} P(\mathbf{x}, \theta_1)$$

Scanned with OKEN Scanner

și

$$\sum_{x \in W} P(x, \theta_1) = \sum_{x \in W \cap W^*} P(x, \theta_1) + \sum_{x \in W \cap \overline{W}^*} P(x, \theta_1)$$

obținem

$$\delta = \sum_{x \in W^* \cap \overline{W}} P(x, \theta_1) - \sum_{x \in W \cap \overline{W}^*} P(x, \theta_1)$$

Din (2.4) și (2.5) obținem

$$\delta \geqslant k\left[\sum_{x \in W^* \cap \overline{W}} P(x, \theta_0) - \sum_{x \in W \cap \overline{W}^*} P(x, \theta_0)\right]$$

sau

$$\delta \geqslant k\left[\sum_{x \in W^*} P(x, \theta_0) - \sum_{x \in W} P(x, \theta_0)\right]$$

Ținînd seama că pentru orice submulțime W a lui $\mathfrak{R}$ are loc (1.1), obținem $\delta \geqslant k(\alpha - \alpha) = 0$, ceea ce demonstrează lema.

Deci lema stabilește ca regiune critică acea mulțime a valorilor $(x_1, \ldots, x_n)$ pentru care

$$\frac{P(x, \theta_1)}{P(x, \theta_0)} = \frac{\prod_1^n p(x_i, \theta_1)}{\prod_1^n p(x_i, \theta_0)} \geqslant k \tag{2.9}$$

În cazul în care avem o singură observație, regiunea critică este dată de

$$\frac{p(x, \theta_1)}{p(x, \theta_0)} \geqslant k \tag{2.9'}$$

Se observă că în acest caz regiunea critică este o parte a dreptei reale.

§ 3. Teste unilaterale privind media repartiției normale cu dispersia cunoscută

Cînd ipoteza alternativă nu este o ipoteză simplă (adică $\theta = \theta_1$) ci o ipoteză compusă (adică $\theta > \theta_0$ sau $\theta \neq \theta_0$), nu există întotdeauna un test care să fie mai puternic ca oricare alt test

Scanned with OKEN Scanner

pentru toate alternativele posibile, sau cu alte cuvinte nu există întotdeauna un cel mai puternic test.

În cazul în care dorim să verificăm ipoteza nulă, conform căreia o populație normală are media $\mu = \mu_0$, față de ipoteza alternativă $\mu > \mu_0$ există un cel mai puternic test.

Considerăm o valoare particulară $\mu_1 > \mu_0$. Pentru această valoare, din lema Neyman-Pearson, cel mai bun test este acela pentru care domeniul lui x determinînd regiunea critică satisface inegalitatea

$$\frac{p(x, \mu_1, \sigma^2)}{p(x, \mu_0, \sigma^2)} > k \tag{3.1}$$

sau, ținînd seama că

$$p(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

membrul stîng al inegalității (3.1) se scrie

$$\frac{p(x, \mu_1, \sigma^2)}{p(x, \mu_0, \sigma^2)} = [\exp x(\mu_1 - \mu_0)/\sigma^2][\exp -(\mu_1^2 - \mu_0^2)/2\sigma^2]. \tag{3.2}$$

Deoarece μ_0, μ_1 sînt specificate și $\mu_1 > \mu_0$ iar σ^2 cunoscută, $\frac{p(x, \mu_1, \sigma^2)}{p(x, \mu_0, \sigma^2)}$ crește odată cu x. Astfel pentru a obține valorile x pentru care este satisfăcută (3.1), pornim de la $x = +\infty$ spre stînga, pînă cînd aria de sub $p(x, \mu_0, \sigma^2)$, de la dreapta lui x, este egală cu pragul de semnificație α. Cea mai mare valoare a lui x din regiunea de acceptare x_c se obține prin transformarea lui (3.1) în egalitate. În acest caz din (3.2) obținem

$$[\exp x(\mu_1 - \mu_0)/\sigma^2][\exp -(\mu_1^2 - \mu_0^2)/2\sigma^2] = k,$$

sau

$$\frac{x(\mu_1 - \mu_0)}{\sigma^2} - \frac{\mu_1^2 - \mu_0^2}{2\sigma^2} = \ln k, \tag{3.3}$$

care rezolvată în raport cu x dă

$$x_c = \frac{2\sigma^2 \ln k + (\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)}. \tag{3.4}$$

Scanned with OKEN Scanner

Regiunea critică a testului este dată de mulțimea acelor x care sînt mai mari decît x_c. În (3.4), μ_0, μ_1 și σ^2 sînt constante și k poate fi determinat astfel încît

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{x_c}^{\infty}\exp\left[-\frac{(x-\mu_0)^2}{2\sigma^2}\right]dx=\alpha. \tag{3.5}$$

Această ecuație determină valoarea lui x_c, și deci regiunea critică $x > x_c$. În acest mod, problema determinării regiunii critice este rezolvată fără cunoașterea lui k.

Din (3.5) obținem

$$P(x<x_c)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_c}\exp\left[-\frac{(v-\mu_0)^2}{2\sigma^2}\right]dv=$$

$$=1-\alpha=\Phi\left(\frac{x_c-\mu_0}{\sigma}\right)^{*)}$$

de unde

$$\frac{x_c-\mu_0}{\sigma}=u_{1-\alpha}$$

sau

$$x_c=\mu_0+\sigma u_{1-\alpha} \tag{3.6}$$

ceea ce arată că x_c nu depinde de μ_1.

Puterea acestui test este dată de probabilitatea de a respinge ipoteza nulă: $\mu=\mu_0$ cînd $\mu=\mu_1$, adică probabilitatea ca observația x să varieze în regiunea critică $(x > x_c)$. Puterea, cînd $\mu=\mu_1$, este

$$\pi(\mu_1)=P(x>x_c\,|\,\mu=\mu_1)=\frac{1}{\sigma\sqrt{2\pi}}\int_{x_c}^{\infty}\exp\left[-\left(\frac{x-\mu_1}{\sqrt{2}\sigma}\right)^2\right]dx=$$

(3.7)

$$=1-\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x_c}\exp\left[-\left(\frac{x-\mu_1}{\sqrt{2}\sigma}\right)^2\right]dx=1-\Phi\left(\frac{x_c-\mu_1}{\sigma}\right).$$

---

*) Valorile funcției $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt$ sînt date în tabela 1.

Scanned with OKEN Scanner

Dacă în (3.7) ţinem seama de (3.6) şi de egalităţile

$$1 - \Phi(v) = \Phi(-v), \quad u_\alpha = -u_{1-\alpha},$$

obţinem

$$\pi(\mu_1) = 1 - \Phi\left(u_{1-\alpha} + \frac{\mu_0 - \mu_1}{\sigma}\right) = \Phi\left(-u_{1-\alpha} - \frac{\mu_0 - \mu_1}{\sigma}\right) =$$

(3.8)

$$= \Phi\left(u_\alpha + \frac{\mu_1 - \mu_0}{\sigma}\right).$$

Am făcut verificarea pe baza unei singure observaţii. În cele ce urmează să presupunem că $x_1, x_2, \ldots, x_n$ este o selecţie dintr-o populaţie $N(\mu, \sigma^2)$ unde μ este necunoscută, iar σ^2 cunoscută. Ne propunem să verificăm ipoteza $H_0 : \mu = \mu_0$ faţă de alternativa $\mu > \mu_0$ la pragul de semnificaţie α. Avem

$$\frac{\prod_1^n p(x_i, \mu_1, \sigma^2)}{\prod_1^n p(x_i, \mu_0, \sigma^2)} = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu_1)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu_0)^2\right]} =$$

(3.9)

$$= \exp\left[\bar{x}\,\frac{\mu_1 - \mu_0}{\sigma^2/n}\right] \exp\left(-\frac{\mu_1^2 - \mu_0^2}{2\sigma^2/n}\right)$$

relaţie similară cu (3.2) în care x este înlocuit prin $\bar{x}$, iar σ^2 prin σ^2/n.

Regiunea critică, analog cu (3.6), va fi

(3.10)
$$\bar{x} > \mu_0 + \frac{\sigma u_{1-\alpha}}{\sqrt{n}}$$

şi puterea, analog cu (3.8), va fi

(3.11)
$$\pi(\mu_1) = \Phi\left(u_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right).$$

(3.11) se poate folosi pentru a determina numărul necesar de observaţii pentru a obţine o putere $(1 - \beta)$ dată, în ipoteza că σ este cunoscută. În acest scop vom egala (3.11) cu $1 - \beta$.

Scanned with OKEN Scanner

Avem

$$\Phi\left(u_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) = 1 - \beta$$

sau

$$u_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} = u_{1-\beta}.$$

Ţinînd seama că $u_{1-\alpha} = -u_\alpha$, şi rezolvînd în raport cu n, obţinem

$$n = (u_{1-\beta} + u_{1-\alpha})^2 \left(\frac{\sigma}{\mu_1 - \mu_0}\right)^2. \tag{3.12}$$

Cînd $H_1 : \mu < \mu_0$

— regiunea critică este $x < x_c = \mu_0 + \sigma u_\alpha$ pentru o singură observaţie,

$\bar{x} < x_c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_\alpha$ pentru media a n observaţii,

— puterea este

$$\pi(\mu_1) = \Phi\left(u_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$ pentru o selecţie de volum n.

§ 4. Verificarea ipotezelor compuse

Să presupunem că avem o repartiţie ce depinde de k parametri $\theta_1, \theta_2, \ldots, \theta_k$ şi că ipoteza nulă este compusă, parametrii variind în regiunea ω. Funcţia

$$P(x_1, \ldots, x_n;\ \theta_1, \theta_2, \ldots, \theta_k) = \prod_1^n p(x_i, \theta_1, \theta_2, \ldots, \theta_k) \tag{4.1}$$

poartă numele de *funcţie de verosimilitate*. Prin metoda verosimilităţii maxime putem găsi estimatorii de maximă verosimilitate pentru parametrii θ şi prin definiţie aceştia maximizează funcţia de verosimilitate pe care s-o notăm $(P)_\omega$.

Să presupunem că parametrii pot varia în spaţiul Ω permis atît de ipoteza nulă cît şi de ipoteza alternativă; găsim şi în acest caz estimatorii de maximă verosimilitate şi deci maximul $(P)_\Omega$ în ambele ipoteze. În acest caz extensia naturală a rela-

ției (2.9) de la ipoteze simple, la ipoteze compuse este raportul $(P)_\Omega/(P)_\omega$. Este mai comod să se folosească inversul acestui raport notat prin

$$l = \frac{(P)_\omega}{(P)_\Omega}; \tag{4.2}$$

l poartă numele de *raport de verosimilitate.*
l este pozitiv deoarece este raportul a două funcții de verosimilitate care sînt pozitive. l nu poate fi mai mare decît 1 deoarece $(P)_\omega$, maximul funcției de verosimilitate cu parametrii variind în ω, nu poate fi mai mare decît $(P)_\Omega$ maximul funcției de verosimilitate cu parametrii variind în Ω. Prin urmare

$$0 \leqslant l \leqslant 1.$$

O valoare mică a lui l indică că verosimilitatea calculată folosind ω corespunzătoare la ipoteza nulă, conduce la respingerea ipotezei nule. O valoare a lui l apropiată de 1 indică că ipoteza nulă este plauzibilă și o vom accepta. Se definește regiunea critică ca

$$0 < l < c \tag{4.3}$$

unde c este o constantă aleasă astfel încît pragul de semnificație al testului este α.

§ 5. Test bilateral privind media populației normale cu dispersia cunoscută

Fie $x_1, \ldots, x_n$ o selecție de volum n dintr-o populație normală de parametri μ și σ^2 (σ^2 cunoscută). Ne propunem să verificăm ipoteza $\mu = \mu_0$ față de alternativa $\mu = \mu_1$, $\mu_1 \neq \mu_0$.

Funcția de verosimilitate este

$$P = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu)^2\right],$$

iar estimația de maximă verosimilitate pentru μ este

$$\hat{\mu} = \frac{1}{n}\sum_1^n x_i = \bar{x}.$$

Scanned with OKEN Scanner

Obținem

$$(P)_\Omega = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \bar{x})^2\right]$$

și ținînd seama că în ipoteza nulă nu sînt parametri care să varieze (σ^2 cunoscută și fixată, $\mu = \mu_0$ dat),

$$(P)_\omega = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu_0)^2\right].$$

Urmează că

$$l = \frac{\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu_0)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \bar{x})^2\right]}$$

sau ținînd seama că

$$\sum_1^n (x_i - \mu_0)^2 = \sum_1^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2,$$

(5.1)

$$l = \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2\right].$$

Marginea regiunii critice $\bar{x}_c$ este determinată de

$$\exp\left[-\frac{n}{2\sigma^2}(\bar{x}_c - \mu_0)^2\right] = c$$

sau

(5.2) $$\bar{x}_c = \mu_0 - \sqrt{-\ln c}\,\sqrt{2}\,\frac{\sigma}{\sqrt{n}} \quad \text{și} \quad \mu_0 + \sqrt{-\ln c}\,\sqrt{2}\,\frac{\sigma}{\sqrt{n}}.$$

Deoarece c este mai mic decît 1, $\ln c$ este negativ și x_c este de forma

$$\bar{x}_c = \mu_0 + k\frac{\sigma}{\sqrt{n}}$$

unde $k = \sqrt{-2\ln c}$.

Scanned with OKEN Scanner

Din (5.1) se observă că dacă $\bar{x}$ este îndepărtată de μ_0, atunci $(\bar{x} - \mu_0)^2$ va avea o valoare mare, iar l o valoare mică. În acest caz valorile mari ale modulului $|\bar{x} - \mu_0|$ constituie un semnal pentru respingerea ipotezei nule. Regiunea critică este partea dreptei reale de la stînga lui $\mu_0 - k\frac{\sigma}{\sqrt{n}}$ plus partea dreptei reale de la dreapta lui $\mu_0 + k\frac{\sigma}{\sqrt{n}}$.

Aria totală de sub curbă, în regiunea critică, este α și cum curba normală este simetrică față de medie, urmează că aria fiecărei cozi este $\frac{\alpha}{2}$.

Regiunea critică va fi

$$\bar{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} \text{ și } \bar{x} > \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \tag{5.4}$$

și ținînd seama că $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$

$$\bar{x} < \mu_0 + \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} \quad \text{se scrie} \quad \bar{x} < \mu_0 - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$$

Se respinge ipoteza H_0 dacă

$$\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > u_{1-\frac{\alpha}{2}}. \tag{5.5}$$

Puterea acestui test, μ fiind egal cu μ_1, este

$$\pi(\mu_1) = P\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -u_{1-\frac{\alpha}{2}}\right) + P\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > u_{1-\frac{\alpha}{2}}\right) =$$

$$= P\left(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} < -u_{1-\frac{\alpha}{2}}\right) +$$

$$+ P\left(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} > u_{1-\frac{\alpha}{2}}\right) =$$

$$= P\left(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} < u_{\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + 1 - P\left(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} < -u_{\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right).$$

Scanned with OKEN Scanner

Dacă $\mu = \mu_1$ variabila aleatoare $\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}}$ este repartizată $N(0, 1)$, ceea ce ne permite să scriem (5.6) sub forma

$$\pi(\mu_1) = \Phi\left(u_{\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + 1 - \Phi\left(-u_{\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) =$$

(5.7)

$$= \Phi\left(u_{\frac{\alpha}{2}} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(u_{\frac{\alpha}{2}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right).$$

Dacă $\mu_1 = \mu_0$ puterea coincide cu pragul de semnificație. Formula (5.7) se poate folosi pentru determinarea numărului de observații necesar pentru o putere specificată $(1 - \beta)$, presupunînd că σ este cunoscut. Ecuația (5.7) poate fi rezolvată în raport cu n numai prin iterații.

Pentru orice calcul în care $\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} > 0{,}5$, (5.7) se scrie

$$\pi(\mu_1) \simeq \Phi\left(u_{\frac{\alpha}{2}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

și punînd $\pi(\mu_1) = 1 - \beta$, obținem

$$u_{\frac{\alpha}{2}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \simeq u_{1-\beta}$$

de unde

(5.8)
$$n \simeq \left(u_{1-\beta} + u_{1-\frac{\alpha}{2}}\right)\left(\frac{\sigma}{\mu_1 - \mu_0}\right)^2.$$

§ 6. Test bilateral privind media unei populații normale cu dispersia necunoscută

Ne propunem pe baza unei selecții $x_1, x_2, \ldots, x_n$ dintr-o populație normală să verificăm ipoteza

$$H_0: \ \mu = \mu_0, \ \sigma^2 \text{ necunoscută}$$

față de ipoteza alternativă

$$H_1: \ \mu_1 \neq \mu_0.$$

Scanned with OKEN Scanner

În acest caz funcţia de verosimilitate este

$$P = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_1^n (x_i - \mu)^2\right]$$

şi

$$\ln P = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_1^n (x_i - \mu)^2;$$

urmează că

$$\frac{\partial \ln P}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \sum_1^n \frac{x_i}{n}$$

$$\frac{\partial \ln P}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \sum_1^n \frac{(x_i - \bar{x})^2}{n}$$

de unde

$$(P)_\Omega = \left(\frac{n}{2\pi\sum_1^n (x_i - \bar{x})^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}. \tag{6.1}$$

Estimaţia de maximă verosimilitate pentru σ^2, în cazul ipotezei nule ($\mu = \mu_0$), se obţine din ecuaţia

$$\frac{d \ln P}{d\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_1^n (x_i - \mu_0)^2 = 0.$$

Avem

$$\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_i - \mu_0)^2.$$

Urmează că

$$(P)_\omega = \left(\frac{n}{2\pi\sum_1^n (x_i - \mu_0)^2}\right)^{\frac{n}{2}} e^{-\frac{n}{2}}. \tag{6.2}$$

Scanned with OKEN Scanner

(4.2) se scrie

$$l = \left(\frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n (x_i - \mu_0)^2} \right)^{\frac{n}{2}}.$$

Deoarece

$$\sum_1^n (x_i - \mu_0)^2 = \sum_1^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$$

urmează că

$$\frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n (x_i - \mu_0)^2} = \frac{1}{1 + \dfrac{n(\bar{x} - \mu_0)^2}{\sum_1^n (x_i - \bar{x})^2}}$$

sau punînd

(6.3) $$t = \frac{\bar{x} - \mu_0}{\sqrt{\dfrac{\sum_1^n (x_i - \bar{x})^2}{n-1}} \Big/ \sqrt{n}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

obținem

$$l = \left(\frac{1}{1 + \dfrac{t^2}{n-1}} \right)^{\frac{n}{2}}.$$

Deoarece orice funcție monotonă de l poate fi folosită ca însăși l în definirea regiunii critice, vom folosi rădăcina de ordinul n a lui l^2

(6.4) $$l^{\frac{2}{n}} = \frac{1}{1 + \dfrac{t^2}{n-1}}.$$

Scanned with OKEN Scanner

Reamintim că respingem ipoteza nulă pentru valori mari ale lui l, sau ținînd seama de (6.4), pentru valori mari ale lui $|t|$. Astfel t definită de (5.3) poate fi folosită ca o statistică pentru verificarea ipotezei $H_0 : \mu = \mu_0$ față de ipoteza alternativă $H_1 : \mu \neq \mu_0$, σ^2 fiind necunoscut.

§ 7. Compararea a două medii

Fie $x_{11}, \ldots, x_{1n_1}$ și $x_{21}, \ldots, x_{2n_2}$ două selecții de volum n_1, respectiv n_2 din populații $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$, σ_1^2 și σ_2^2 fiind cunoscute.

Avem

$$\bar{x}_1 = \frac{1}{n_1}\sum_1^{n_1} x_{1i}, \quad \bar{x}_2 = \frac{1}{n_2}\sum_1^{n_2} x_{2i},$$

de unde

$$M(\bar{x}_1) = \mu_1, \quad M(\bar{x}_2) = \mu_2,$$

$$D^2(\bar{x}_1) = \frac{\sigma_1^2}{n_1}, \quad D^2(\bar{x}_2) = \frac{\sigma_2^2}{n_2}, \quad \operatorname{cov}(\bar{x}_1, \bar{x}_2) = 0.$$

Prin urmare

$$M(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2 = \delta,$$

$$D^2(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

ceea ce ne permite să afirmăm că variabila aleatoare

$$(7.1) \qquad u = \frac{(\bar{x}_1 - \bar{x}_2) - M(\bar{x}_1 - \bar{x}_2)}{D(\bar{x}_1 - \bar{x}_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

este repartizată normal $N(0, 1)$.

La fel ca în paragrafele anterioare putem găsi teste pentru verificarea ipotezei conform căreia media de selecție $\bar{x}$ cu dispersia $\frac{\sigma^2}{n}$ provine dintr-o populație normală de medie μ_0; de asemenea pentru verificarea ipotezei conform căreia $\bar{x}_1 - \bar{x}_2$ provine dintr-o populație normală de medie δ_0.

Scanned with OKEN Scanner

Pentru verificarea ipotezei $\delta = \delta_1 > \delta_0$, (3.6) dă ca regiune critică

$$\bar{x}_1 - \bar{x}_2 > \delta_0 + u_{1-\alpha} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{7.2}$$

și pentru verificarea ipotezei bilaterale $\delta = \delta_1 \neq \delta_0$, (5.4) dă ca regiune critică

$$\bar{x}_1 - \bar{x}_2 < \delta_0 + u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{și} \quad \bar{x}_1 - \bar{x}_2 > \delta_0 + \tag{7.3}$$

$$+ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\frac{\alpha}{2}}.$$

Scanned with OKEN Scanner

CAPITOLUL

2

VERIFICĂRI DE IPOTEZE PRIVIND PARAMETRII REPARTIȚIILOR BINOMIALĂ, HIPERGEOMETRICĂ ȘI POISSON

§ 1. Repartiția binomială

Dacă un eveniment A are probabilitatea de realizare θ în fiecare probă, atunci probabilitatea ca în n probe evenimentul A să se realizeze de x ori este

$$b(n, x; \theta) = C_n^x \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \ldots, n. \tag{1.1}$$

Pentru demonstrația acestei afirmații a se vedea Ciucu, G., Craiu, V. [4], Cap. XII, teorema 1.

Se observă că $b(n, x; \theta)$ poate fi considerată ca termenul general al dezvoltării binomului lui Newton: $[\theta + (1 - \theta)]^n$.

Mulțimea perechilor ordonate

$$(x, b(n, x; \theta)), \quad x = 0, 1, \ldots, n \tag{1.2}$$

constituie repartiția binomială.

Media și dispersia repartiției binomiale (1.2) sînt $n\theta$, respectiv $n\theta(1 - \theta)$.

§ 2. Legătura dintre repartiția binomială și repartiția normală

Arătăm că repartiția binomială poate fi aproximată cu o repartiție normală cu media $n\theta$ și dispersia $n\theta\,(1 - \theta)$.

În acest scop folosim formula lui Stirling

$$n! \simeq n^n e^{-n} \sqrt{2\pi n}. \tag{2.1}$$

Avem

$$C_n^x = \frac{1}{\sqrt{2\pi n}} \left(\frac{n}{x}\right)^{x+\frac{1}{2}} \left(\frac{n}{n-x}\right)^{n-x+\frac{1}{2}}$$

și

$$\theta^x(1-\theta)^{n-x} = \frac{\theta^{x+\frac{1}{2}}}{|\theta} \cdot \frac{(1-\theta)^{n-x+\frac{1}{2}}}{\sqrt{1-\theta}},$$

de unde

$$(2.2) \qquad b(n, x; \theta) \simeq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n\theta(1-\theta)}} \frac{1}{\left(\frac{x}{n\theta}\right)^{x+\frac{1}{2}}} \frac{1}{\left(\frac{n-x}{n(1-\theta)}\right)^{n-x+\frac{1}{2}}}.$$

(2.2) se mai poate scrie

$$(2.3) \quad \ln b(n, x; \theta) = -\ln \sqrt{2\pi n\theta(1-\theta)} - \left(x+\frac{1}{2}\right)\ln\frac{x}{n\theta} - \left(n-x+\frac{1}{2}\right)\ln\frac{n-x}{n(1-\theta)}.$$

Dacă notăm variabila redusă prin z,

$$(2.4) \qquad z = \frac{x - M(x)}{D(x)} = \frac{x-n\theta}{n\theta(1-\theta)},$$

ținem seama că

$$\frac{x}{n\theta} = 1 + z\sqrt{\frac{1-\theta}{n\theta}},$$

$$\frac{n-z}{n(1-\theta)} = 1 - z\sqrt{\frac{\theta}{n(1-\theta)}},$$

și

$\ln(1+y) = y - \frac{y^2}{2} + \frac{y^3}{3}$ ... (convergență pentru $|y| < 1$) atunci (2.3) se scrie

$$\ln b(n, x; \theta) \simeq -\ln\sqrt{2\pi n\theta(1-\theta)} - \left(n\theta + z\sqrt{n\theta(1-\theta)} + \frac{1}{2}\right)\left(z\sqrt{\frac{1-\theta}{n\theta}} - z^2\frac{1-\theta}{2n\theta} + \ldots\right) - \left(n(1-\theta) - z\sqrt{n\theta(1-\theta)} + \frac{1}{2}\right)\left(-z\sqrt{\frac{\theta}{n(1-\theta)}} - z^2\frac{\theta}{2n(1-\theta)} - \ldots\right)$$

sau

$$\ln b(n, x; \theta) \simeq -\ln \sqrt{2\pi n\theta(1-\theta)} - \frac{z^2}{2}, \tag{2.5}$$

termenii de ordin superior tinzînd la zero.

Ţinînd seama de (2.4) în (2.5) se obţine

$$b(n, x; \theta) \simeq \frac{1}{\sqrt{2\pi n\theta(1-\theta)}} \exp\left[-\frac{(x - n\theta)^2}{2n\theta(1-\theta)}\right], \tag{2.6}$$

care este tocmai densitatea de repartiţie corespunzătoare unei repartiţii $N(n\theta, n\theta(1-\theta))$.

Observaţii. 1°. S-a aproximat o repartiţie discretă printr-o repartiţie continuă.

2°. În cazul în care x are o repartiţie discretă probabilităţile se atribuie întregilor, iar în cazul aproximaţiei continue probabilitatea corespunzătoare este distribuită pe aria unui dreptunghi avînd baza egală cu unitatea, la distanţă $\frac{1}{2}$ de o parte şi de alta a întregului.

Integrala

$$\int_{-\infty}^{x_1} b(n, x; \theta)\, dx$$

nu conţine aria de la x_1 la $x_1 + \frac{1}{2}$, care de fapt aparţine dreptunghiului avînd mijlocul bazei în x_1. Prin urmare va trebui să schimbăm limita superioară de integrare la $x_1 + \frac{1}{2}$.

$$P(x \leqslant x_1) \simeq \int_{-\infty}^{x_1+\frac{1}{2}} b(n, x; \theta)\, dx = \Phi\left(\frac{x_1 + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right). \tag{2.7}$$

Pe de altă parte dacă dorim să determinăm probabilitatea că $X < x_1$, integrala

$$\int_{-\infty}^{x_1} b(n, x; \theta)\, dx$$

Scanned with OKEN Scanner

conține aria de la $x_1 - \frac{1}{2}$ la x_1, ceea ce nu trebuie. Prin urmare, vom schimba limita superioară de integrare la $x_1 - \frac{1}{2}$

$$(2.8) \qquad P(x < x_1) = \int_{-\infty}^{x_1 - \frac{1}{2}} b(n, x; \theta)\, dx = \Phi\left(\frac{x_1 - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right).$$

Constanta $\frac{1}{2}$ care este adunată sau scăzută la limita superioară de integrare este cunoscută sub numele de *corecția de continuitate.*

Deoarece

$$P(x = x_1) = P(x \leqslant x_1) - P(x < x_1)$$

ținînd seama de (2.7) și (2.8) obținem

$$(2.9) \qquad P(x = x_1) = \Phi\left(\frac{x_1 + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{x_1 - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right).$$

Adeseori prezintă importanță frecvența relativă $\frac{x}{n}$ de realizare a evenimentului A. În acest caz (2.6) devine

$$(2.6') \qquad b^*\left(n, \frac{x}{n}; \theta\right) = \frac{1}{\sqrt{2\pi}\,\sqrt{\theta(1-\theta)/n}} \exp\left[-\frac{\left(\frac{x}{n} - \theta\right)^2}{2\theta(1-\theta)/n}\right]$$

iar echivalentele lui (2.7) și (2.8)

$$(2.7') \qquad P\left(\frac{x}{n} \leqslant a\right) \simeq \Phi\left(\frac{a + \frac{1}{2n} - \theta}{\sqrt{\theta(1-\theta)/n}}\right),$$

$$(2.8') \qquad P\left(\frac{x}{n} < a\right) \simeq \Phi\left(\frac{a - \frac{1}{2n} - \theta}{\sqrt{\theta(1-\theta)/n}}\right).$$

Scanned with OKEN Scanner

§ 3. Verificarea ipotezelor privind parametrii repartiției binomiale cu ajutorul aproximației normale

Presupunem că într-o selecție de volum n, dintr-o populație binomială, se găsesc x defecte și că dorim să verificăm ipoteza conform căreia proporția defectelor θ în populație are valoarea particulară θ_0.

a) Ipoteza alternativă este $\theta = \theta_1 > \theta_0$.

Cea mai rațională regiune critică este dată de acea parte a axei x-lor care corespunde la o suprafață de arie α în coadă superioară a repartiției lui x (fig. 2.1) cînd este adevărată ipo-

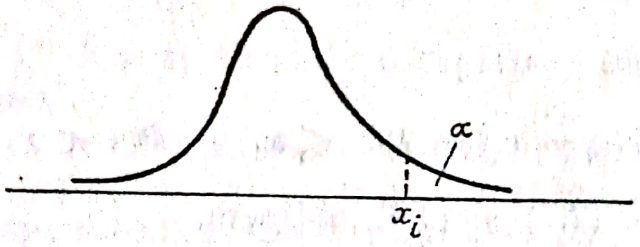


Fig. 2.1

teza nulă. Dacă x_i reprezintă coordonata x corespunzătoare marginii inferioare a regiunii critice, atunci regiunea de acceptare se va întinde pînă la $(x_i - 1)$. Cerem ca probabilitatea de cădere în regiunea de acceptare, atunci cînd ipoteza H_0 este adevărată, să fie $1 - \alpha$. Ținînd seama de (2.8), x_i este determinat din ecuația

$$P(x < x_i) \simeq \Phi\left(\frac{x_i - \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\right) = 1 - \alpha \tag{3.1}$$

de unde

$$\frac{x_i - \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} = u_{1-\alpha},$$

care rezolvată în raport cu x_i conduce la

$$x_i = n\theta_0 + \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{1-\alpha} \tag{3.2}$$

valoare care în general nu este un întreg.

Scanned with OKEN Scanner

Vom accepta ipoteza nulă dacă $x < x_t$ și o respingem dacă $x \geqslant x_t$. Dacă folosim ca margine inferioară pentru regiunea critică, primul întreg mai mic decît x_t, x', atunci regiunea critică va avea un α mai mare decît pragul de semnificație dorit de noi. Dacă x' este primul întreg mai mare decît x_t, atunci regiunea critică va fi mai mică. Valorile numerice aproximative se obțin introducînd x' și (3.1) și rezolvînd pentru valoarea corespunzătoare a lui α, α'.

Dacă x_1 este numărul de defecte, cerem ca

$$P(x \geqslant x_1) = 1 - P(x < x_1) \simeq 1 - \Phi\left(\frac{x_1 - \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\right) \simeq P$$

sau

$$\Phi\left(\frac{x_1 - \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\right) \simeq 1 - P, \tag{3.3}$$

care poate fi rezolvată în raport cu P.

Puterea se calculează folosind (2.8) și (3.2). Avem:

$$\pi(\theta_1) = P(x \geqslant x_t \mid \theta = \theta_1) \simeq 1 -$$

$$- \Phi\left(\frac{n\theta_0 + \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{1-\alpha} - \frac{1}{2} - n\theta_1}{\sqrt{n\theta_1(1-\theta_1)}}\right) =$$

$$= 1 - \Phi\left(\sqrt{n}\,\frac{\theta_0 - \theta_1}{\sqrt{\theta_1(1-\theta_1)}} + \sqrt{\frac{\theta_0(1-\theta_0)}{\theta_1(1-\theta_1)}}\, u_{1-\alpha}\right). \tag{3.4}$$

Această formulă poate fi folosită pentru determinarea volumului selecției necesar pentru o putere $(1 - \beta)$ de respingere a ipotezei nule la pragul de semnificație α cînd $\theta_1 - \theta_0$ are o valoare specifică.

(3.4) egalat cu $1 - \beta$ conduce la

$$n = \frac{1}{(\theta_1 - \theta_0)^2}\left(u_{1-\beta}\sqrt{\theta_1(1-\theta_1)} + u_{1-\alpha}\sqrt{\theta_0(1-\theta_0)}\right)^2. \tag{3.5}$$

Scanned with OKEN Scanner

b) În cazul în care ipoteza alternativă este $\theta_1 < \theta_0$ cerem ca probabilitatea ca x să cadă în regiunea $\leqslant x_i$, atunci cînd este adevărată ipoteza H_0, să fie α. Din (2.7) obținem

$$P(x \leqslant x_i) \simeq \Phi\left(\frac{x + \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\right) = \alpha$$

de unde

$$x_i \simeq n\theta_0 - \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_\alpha. \tag{3.6}$$

Dacă $x > x_i$, acceptăm ipoteza nulă, iar dacă $x \leqslant x_i$ respingem ipoteza nulă.

Dacă x_1 este numărul de defecte cerem ca

$$P(x \leqslant x_1) \simeq \Phi\left(\frac{x_1 + \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}\right) \simeq P$$

sau

$$\frac{x_1 + \frac{1}{2} - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} = u_P, \tag{3.7}$$

care poate fi rezolvată în raport cu P.

Puterea este

$$\pi(\theta_1) = P(x \leqslant x_i \mid \theta = \theta_1) \simeq$$

$$\simeq \Phi\left(\frac{n\theta_0 - \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_\alpha + \frac{1}{2} - n\theta_1}{\sqrt{n\theta_1(1-\theta_1)}}\right) =$$

$$= \Phi\left(\sqrt{n}\,\frac{\theta_0 - \theta_1}{\sqrt{\theta_1(1-\theta_1)}} + \sqrt{\frac{\theta_0(1-\theta_0)}{\theta_1(1-\theta_1)}}\, u_\alpha\right). \tag{3.8}$$

Scanned with OKEN Scanner

c) În cazul în care ipoteza alternativă este bilaterală: $\theta_1 \neq \theta_0$, regiunea critică este

$$x \leqslant n\theta_0 - \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{\frac{\alpha}{2}} \tag{3.9}$$

și

$$x \geqslant n\theta_0 + \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{1-\frac{\alpha}{2}}. \tag{3.10}$$

În acest caz puterea este dată de

$$\pi(\theta_1) = P\left(x \leqslant n\theta_0 - \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{\frac{\alpha}{2}} \mid \theta = \theta_1\right) + \tag{3.11}$$

$$+ P\left(x \geqslant n\theta_0 + \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{1-\frac{\alpha}{2}} \mid \theta = \theta_1\right) =$$

$$= \Phi\left(\frac{n\theta_0 - \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{\frac{\alpha}{2}} + \frac{1}{2} - n\theta_1}{\sqrt{n\theta_1(1-\theta_1)}}\right) +$$

$$+ 1 - \Phi\left(\frac{n\theta_0 + \frac{1}{2} + \sqrt{n\theta_0(1-\theta_0)}\, u_{1-\frac{\alpha}{2}} - \frac{1}{2} - n\theta_1}{\sqrt{n\theta_1(1-\theta_1)}}\right) =$$

$$= \Phi\left(\sqrt{n}\,\frac{\theta_0 - \theta_1}{\sqrt{\theta_1(1-\theta_1)}} + \sqrt{\frac{\theta_0(1-\theta_0)}{\theta_1(1-\theta_1)}}\, u_{\frac{\alpha}{2}}\right) + \Phi\left(\sqrt{n}\sqrt{\frac{\theta_0(1-\theta_0)}{\theta_1(1-\theta_1)}} -\right.$$

$$\left. - \sqrt{n}\,\frac{\theta_0 - \theta_1}{\sqrt{\theta_1(1-\theta_1)}}\, u_{\frac{\alpha}{2}}\right)$$

§ 4. Transformări unghiulare și alte transformări

Este preferabil să presupunem că dispersia variabilei este constantă. Există însă situații în care dispersia este o funcție de media variabilei. De exemplu dispersia frecvenței relative a evenimentului A: $\left(\frac{x}{n}\right)$ este $\theta(1-\theta)/n$ deci funcție de θ. În aceste

situații, pentru a putea folosi metodele statistice care presupun cunoscută dispersia, vom găsi o funcție de $x(\nu)$, astfel încît dispersia acestei funcții să fie constantă. Aceasta se poate realiza după cum urmează.

Fie $y = \nu(x)$ o funcție de variabila aleatoare x. Folosind dezvoltarea lui Taylor, putem să aproximăm $\nu(x)$ printr-o funcție liniară în vecinătatea lui μ (dacă dispersia lui x este mică)

$$\nu(x) \simeq \nu(\mu) + (x - \mu)\nu'(\mu). \tag{4.1}$$

Aplicînd valoarea medie în ambii membri ai egalității anterioare, obținem

$$M(y) \simeq \nu(\mu), \tag{4.2}$$

de unde

$$y - M(y) = (x - \mu)\,\nu'(\mu)$$

și

$$D^2(y) = M[y - M(y)]^2 = [\nu'(\mu)]^2\, M(x - \mu)^2 = [\nu'(\mu)]^2\, D^2(x). \tag{4.3}$$

În general $D^2(x)$ depinde de anumiți parametri. Dacă $D^2(x)$ poate fi scrisă ca o funcție de $\mu[g(\mu)]$, atunci luînd $\nu'(\mu) = 1/\sqrt{g(\mu)}$ obținem $D^2(y) = 1$. Ecuația

$$\nu'(\mu) = \frac{1}{\sqrt{g(\mu)}} \tag{4.4}$$

este echivalentă cu ecuația

$$\nu(\mu) = \int \frac{d\mu}{\sqrt{g(\mu)}}. \tag{4.5}$$

Exemplu. Pentru frecvența relativă $\frac{x}{n}$, cu $M\left(\frac{x}{n}\right) = \theta$, $D^2\left(\frac{x}{n}\right) = \frac{\theta(1-\theta)}{n}$, avem

$$\nu(\theta) = \int \frac{d\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} = \sqrt{n} \int \frac{d\theta}{\sqrt{\theta(1-\theta)}}.$$

Scanned with OKEN Scanner

Punînd $t = \sqrt{\theta}$, obţinem

$$\nu(\theta) = \sqrt{n}\int \frac{d\theta}{\sqrt{\theta(1-\theta)}} = \sqrt{n}\, 2\int \frac{dt}{\sqrt{1-t^2}} =$$

$$= 2\sqrt{n}\ \arcsin\ t = 2\sqrt{n}\ \arcsin\sqrt{\theta}.$$

Astfel dacă folosim transformarea

$$\nu\left(\frac{x}{n}\right) = 2\sqrt{n}\ \arcsin\sqrt{\frac{x}{n}},$$

dispersia lui $\frac{x}{n}$ va fi 1. Dacă omitem factorul $\sqrt{n}$ şi folosim transformarea

$$\psi\left(\frac{x}{n}\right) = 2\ \arcsin\sqrt{\frac{x}{n}}$$

atunci dispersia noii variabile este $1/n$.

Alte transformări

1. Dacă dispersia este proporţională cu valoarea medie, adică

$$D^2(x) = k\mu,$$

atunci

$$\nu(\mu) = \int \frac{d\mu}{\sqrt{k\mu}} = \frac{2}{\sqrt{k}}\sqrt{\mu}. \tag{4.6}$$

Pentru $k = 1$, ne situăm în cazul repartiţiei Poisson.

2. Dacă $D^2(x) = k^2\mu^2$, atunci

$$\nu(\mu) = \int \frac{d\mu}{\sqrt{k^2\mu^2}} = \frac{1}{k}\ln\ \mu$$

astfel încît ln μ va avea dispersia egală cu constanta k^2, iar lg μ dispersia 0,189 k^2.

Dacă

$$\psi(\mu) = \ln\ \mu,$$

Scanned with OKEN Scanner

atunci

$$y \approx \psi(u) + (x - \mu)\,\psi'(\mu) + \frac{(x-\mu)^2}{2}\,\psi''(\mu),$$

de unde

$$M(y) \simeq \psi(\mu) + \frac{1}{2}\psi''(\mu)\,M(x-\mu)^2 = \ln \mu - \frac{1}{2\mu^2}\,D^2(x).$$

Aceasta ne arată că atunci cînd în seria Taylor includem și termenii de ordinul doi, valoarea medie a variabilei transformate y diferă de (4.2) cu o cantitate care în acest caz este $-\frac{1}{2\mu^2}D^2(x)$.

Observăm că dacă punem în (4.3)

$$\frac{1}{\mu} = \psi'(\mu)$$

obținem

$$D^2(y) = \frac{D^2(x)}{(M(x))^2}.$$

3. Dacă $D^2(x) = k^2\mu^4$, atunci

$$\nu(\mu) = \int \frac{d\mu}{\sqrt{k^2\mu^4}} = \int \frac{d\mu}{k\mu^2} = -\frac{1}{k\mu}.$$

§ 5. Teste privind repartiția binomială folosind transformările unghiulare

Dacă x este numărul de realizări ale evenimentului A în n probe, atunci

$$y = 2 \arcsin \sqrt{\frac{x}{n}};$$

în radiani este aproximativ normal repartizată

$$N\left(2 \arcsin \sqrt{\theta},\ \frac{1}{n}\right).$$

Scanned with OKEN Scanner

Urmează că

$$(5.1)\quad P\left(\frac{x}{n} \leqslant \tau\right) = \int_{-\infty}^{2 \arcsin \sqrt{\tau + 1/2n}} f(y)\, dy = \Phi\left(\frac{2 \arcsin \sqrt{\tau + 1/2n} - \eta}{1/\sqrt{n}}\right)$$

unde $1/2n$ este corecția pentru continuitate și $\eta = 2 \arcsin \sqrt{\theta}$.

Test unilateral. Verificăm ipoteza H_0: $\theta = \theta_0$ față de alternativa H_1: $\theta = \theta_1 < \theta_0$. Regiunea critică se obține egalînd (5.1) cu α și punînd

$\underline{\tau}$ pentru τ și $\eta_0 = 2 \arcsin \sqrt{\theta_0}$ pentru η.

Obținem

$$\Phi\left(\frac{2 \arcsin \sqrt{\underline{\tau} + 1/2n} - 2 \arcsin \sqrt{\theta_0}}{1/\sqrt{n}}\right) = \alpha$$

sau

$$\frac{2 \arcsin \sqrt{\underline{\tau} + 1/2n} - 2 \arcsin \sqrt{\theta_0}}{1/\sqrt{n}} = u_\alpha$$

de unde

$$2 \arcsin \sqrt{\underline{\tau} + \frac{1}{2n}} = 2 \arcsin \sqrt{\theta_0} + \frac{u_\alpha}{\sqrt{n}}$$

care pentru valori specificate ale lui θ_0, α și n determină pe $\underline{\tau}$.

Dacă $\frac{x}{n} \leqslant \underline{\tau}$ respingem ipoteza H_0.

Puterea testului pentru alternativa

a) $\eta_1 > \eta_0$ este $\pi(\eta_1) = \Phi\left(u_\alpha + \frac{\eta_1 - \eta_0}{1/\sqrt{n}}\right)$;

b) $\eta_1 < \eta_0$ este $\pi(\eta_1) = \Phi\left(u_\alpha - \frac{\eta_1 - \eta_0}{1/\sqrt{n}}\right)$;

c) $\eta_1 \neq \eta_0$ este $\pi(\eta_1) = \Phi\left(u_{\frac{\alpha}{2}} - \frac{\eta_1 - \eta_0}{1/\sqrt{n}}\right) +$

$$+ \Phi\left(u_{\frac{\alpha}{2}} + \frac{\eta_1 - \eta_0}{1/\sqrt{n}}\right).$$

Scanned with OKEN Scanner

Volumul selecției n este dat de:

a) pentru testul unilateral de $n = (u_{1-\beta} + u_{1-\alpha})^2 \left(\frac{1}{\eta_1 - \eta_0}\right)^2$;

b) pentru testul bilateral de $n = (u_{1-\beta} + u^{1-\frac{\alpha}{2}})^2 \frac{1}{(\eta_1 - \eta_0)^2}$.

§ 6. Verificarea ipotezelor privind repartiția binomială

Dacă x este numărul de realizări ale evenimentului A în urma a n probe ($P(A) = \theta$ în fiecare probă) atunci

$$(6.1) \qquad P(x \leqslant k) = \sum_{i=0}^{k} C_n^i \theta^i (1-\theta)^{n-i} = 1 - P\left(F < \frac{n-k}{k+1} \cdot \frac{\theta}{1-\theta}\right),$$

unde F este repartizată Snedecor cu $2(k+1)$, respectiv $2(n-k)$ grade de libertate. Avem

$$(6.2) \qquad F_\alpha(n_1, n_2) = 1/F_{1-\alpha}(n_2, n_1).$$

§ 7. Compararea a două frecvențe observate cu ajutorul aproximației normale

Presupunem că din populații infinite cu fracția defectelor θ_1, respectiv θ_2, considerăm selecțiile de volum n_1, respectiv n_2. Fie x_1 numărul de piese defecte din prima selecție și x_2 numărul de piese defecte din a doua selecție.

Ne propunem să verificăm ipoteza nulă conform căreia $\theta_1 = \theta_2$. Dacă este adevărată această ipoteză atunci $p_1 \left(= \frac{x_1}{n_1}\right) - p_2 \left(= \frac{x_2}{n_2}\right)$ are media zero și dispersia

$$D^2(p_1 - p_2) = \frac{\theta(1-\theta)}{n_1} + \frac{\theta(1-\theta)}{n_2} = \theta(1-\theta)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

O estimație naturală pentru θ (să o notăm prin p) este

$$p = \frac{x_1 + x_2}{n_1 + n_2}.$$

Scanned with OKEN Scanner

Dacă ipoteza H_0 este adevărată, atunci

$$\frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{7.1}$$

este repartizată normal redus.

O îmbunătățire a acestei aproximații se obține introducînd corecția de continuitate pentru p_1 și p_2, cu un astfel de semn încît valoarea absolută a numărătorului să scadă. Astfel dacă $p_1 > p_2$, o aproximație a variabilei normale reduse este dată de

$$\frac{\left|\left(p_1 - \frac{1}{2n_1}\right) - \left(p_2 - \frac{1}{2n_2}\right)\right|}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \tag{7.2}$$

În cazul unui test bilateral respingem ipoteza H_0 la pragul de semnificație α dacă (7.2) depășește $u_{1-\frac{\alpha}{2}}$.

Puterea testului și determinarea volumului selecției

Pentru determinarea puterii testului și determinarea volumului selecției folosim transformarea unghiulară

$$y_1 = 2 \arcsin \sqrt{\frac{x_1}{n_1}},$$

$$y_2 = 2 \arcsin \sqrt{\frac{x_2}{n_2}}$$

pentru care $D^2(y_1) = \frac{1}{n_1}$, $D^2(y_2) = \frac{1}{n_2}$ astfel ca

$$D^2(y_1 - y_2) = \frac{1}{n_1} + \frac{1}{n_2}.$$

Fie

$$\eta_1 = 2 \arcsin \sqrt{p_1}, \quad \eta_2 = 2 \arcsin \sqrt{p_2}.$$

Scanned with OKEN Scanner

Puterea testului unilateral ($p_1 > p_2$) (1. 3. 8) este

$$\pi(p_1 - p_2) = \Phi\left(u_\alpha + \frac{\eta_1 - \eta_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) \tag{7.3}$$

iar pentru testul bilateral (1. 5. 7)

$$\pi(p_1 - p_2) = \Phi\left(u_{\frac{\alpha}{2}} - \frac{\eta_1 - \eta_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right) + \Phi\left(u_{\frac{\alpha}{2}} + \frac{\eta_1 - \eta_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right). \tag{7.4}$$

Pentru determinarea volumului selecţiilor va trebui să facem unele presupuneri relative la mărimile lui n_1 şi n_2. Se arată că pentru o putere specificată $(n_1 + n_2)$ este minimă cînd $n_1 = n_2 = n$, astfel că

$$D^2(y_1 - y_2) = \frac{2}{n}.$$

Pentru testul bilateral, avem (1. 5. 8)

$$n = (u_{1-\beta} + u_{1-\frac{\alpha}{2}})^2 \frac{2}{(\eta_1 - \eta_2)^2}$$

iar pentru testul unilateral (1.3.11)

$$n = (u_{1-\beta} + u_{1-\alpha})^2 \frac{2}{(\eta_1 - \eta_2)^2}.$$

§ 8. Testul lui Fisher pentru compararea a două proporţii binomiale

În paragraful anterior am prezentat o metodă pentru compararea a două proporţii binomiale folosind aproximaţia normală. Aproximaţia normală nu este acceptabilă în cazul în care avem de-a face cu numere mici. În acest paragraf vom prezenta un test exact care nu prezintă această limitare.

Considerăm toate rezultatele posibile corespunzătoare la două mulţimi de experienţe independente. Prima mulţime se compune din n_1 experienţe şi s-au înregistrat x_1 succese, iar a doua mulţime se compune din n_2 experienţe şi s-au înregistrat x_2 succese, probabilitatea de succes fiind aceeaşi pentru fiecare mulţime.

Scanned with OKEN Scanner

O pereche particulară de succese este $x_1 = a_1$, $x_2 = a_2$.

	Numărul de succese	Numărul de eşecuri	Numărul de experienţe
Prima mulţime	a_1	$n_1 - a_1$	n_1
a doua mulţime	a_2	$n_2 - a_2$	n_2
Total	$a_1 + a_2$	$(n_1 + n_2) - (a_1 + a_2)$	$n_1 + n_2$

În testul exact al lui Fisher se calculează probabilitatea de a obţine succesele $x_1 = a_1$, $x_2 = a_2$, date fiind două selecţii de volum n_1, respectiv n_2 şi numărul total de succese $(a_1 + a_2)$, în ipoteza că probabilitatea de succes este aceeaşi pentru fiecare mulţime.

Avem

$$(8.1)\quad P(x_1 = a_1,\ x_2 = a_2 \mid x_1 + x_2 = a_1 + a_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{P(E_1)}{P(E_2)}$$

unde E_1 este evenimentul $(x_1 = a_1,\ x_2 = a_2)$,
E_2 este evenimentul $x_1 + x_2 = a_1 + a_2$,
deoarece dacă E_1 se realizează, adică dacă $x_1 = a_1$ şi $x_2 = a_2$, atunci se realizează şi E_2, deoarece nu putem avea $x_1 = a_1$ şi $x_2 = a_2$ fără a avea şi $x_1 + x_2 = a_1 + a_2$.

Deoarece cele două mulţimi de experienţă sînt presupuse independente

$$(8.2)\quad P(E_1) = P(x_1 = a_1 \text{ şi } x_2 = a_2) = P(x_1 = a_1)P(x_2 = a_2) =$$

$$= C_{n_1}^{a_1}\theta^{a_1}(1-\theta)^{n_1-a_1}\, C_{n_2}^{a_2}\theta^{a_2}(1-\theta)^{n_2-a_2} =$$

$$= C_{n_1}^{a_1}C_{n_2}^{a_2}\theta^{a_1+a_2}(1-\theta)^{(n_1+n_2)-(a_1+a_2)},$$

$$(8.3)\quad P(E_2) = P(x_1 + x_2 = a_1 + a_2) = C_{n_1+n_2}^{a_1+a_2}\theta^{a_1+a_2}(1-\theta)^{(n_1+n_2)-(a_1+a_2)}.$$

Scanned with OKEN Scanner

Se observă că pentru determinarea probabilității evenimentului E_2 am presupus că într-o selecție de volum $n_1 + n_2$, în care probabilitatea succesului este θ, am avut $a_1 + a_2$ succese.

Ținînd seama în (8.1) de (8.2) și (8.3) obținem

$$P(x_1 = a_1 \text{ și } x_2 = a_2 \mid x_1 + x_2 = a_1 + a_2) = \frac{C_{n_1}^{a_1} C_{n_2}^{a_2}}{C_{n_1+n_2}^{a_1+a_2}} = \tag{8.4}$$

$$= \frac{n_1! n_2! (a_1 + a_2)! (n_1 + n_2 - a_1 - a_2)!}{a_1! (n_1 - a_1)! a_2! (n_2 - a_2)! (n_1 + n_2)!}$$

Aceasta este probabilitatea de a obține tabela observată, date fiind frecvențele marginale, în ipoteza conform căreia selecțiile sînt independente și provin din populația cu același θ. Pentru un test de semnificație, prezintă importanță nu numai această probabilitate, ci și suma probabilităților rezultatelor posibile extreme în aceeași direcție (suma cozii repartiției).

§ 9. Repartiția hipergeometrică

Presupunem că dintr-o populație formată din N elemente, dintre care M sînt defecte, se extrag fără întoarcere n elemente. Se pune problema determinării probabilității ca dintre aceste n elemente m să fie defecte.

N elemente în grupe de cîte n se pot considera în C_N^n moduri, care reprezintă numărul total de cazuri. Deoarece cele m elemente defecte din selecție provin din cele M elemente defecte ale populației, selecția fiind compusă din n elemente, celelalte $n-m$ elemente fără defecte ale selecției vor proveni din cele $N-M$ elemente fără defect ale populației. Prin urmare probabilitatea căutată este dată

$$\frac{\text{numărul modurilor de obținere a selecțiilor de volum } n \text{ conținînd } m \text{ defecte}}{\text{numărul modurilor de obținere a selecțiilor de volum } n} = \tag{9.1}$$

$$= C_M^m C_{N-M}^{n-m} / C_N^n,$$

Scanned with OKEN Scanner

unde

$$0 \leqslant m \leqslant n, \quad 0 \leqslant m \leqslant M, \quad 0 \leqslant n - m \leqslant N - M.$$

Repartiția avînd funcția de frecvență dată de (9.1) poartă numele de *repartiția hipergeometrică.*

§ 10. Legătura dintre repartiția hipergeometrică și repartiția binomială

Presupunînd că N tinde către infinit, $\frac{M}{N}$ fiind constant și egal cu θ, putem scrie

$$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{\dfrac{M!}{m!(M-m)!}\,\dfrac{(N-M)!}{(n-m)!(N-M-n+m)!}}{\dfrac{N!}{n!(N-n)!}} =$$

(10.1)

$$= \frac{n!}{m!(n-m)!}\,\theta^m(1-\theta)^{n-m}.$$

Era de așteptat un asemenea rezultat deoarece dacă N are valori foarte mari, extragerile succesive fără înlocuire nu modifică structura populației, ele fiind independente. Repartiția binomială presupune că extracțiile succesive sînt independente. În situațiile reale, populația nu este infinită. În practică se consideră că repartiția binomială este o aproximație satisfăcătoare pentru repartiția hipergeometrică atunci cînd volumul selecției este mai mic decît 10% din volumul populației.

§ 11. Testul exact al lui Fisher ca o ilustrare a repartiției hipergeometrice

Considerăm o tabelă 2×2 de următoarea formă:

	Defecte	Fără defect	Total
Prima selecție	a	b	$a+b$
A doua selecție	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

Scanned with OKEN Scanner

Testul exact al lui Fisher operează asupra observațiilor marginale, astfel încît din acest punct de vedere avem o populație de volum $N = a + b + c + d$ ce conține $M = a + c$ elemente defecte. Putem privi prima selecție ca o selecție de volum $n = a + b$ din această populație, și în această selecție am observat $m = a$ elemente defecte. Înlocuind în (9.1), obținem

$$(11.1) \qquad \frac{C_{a+c}^{a} C_{b+d}^{b}}{C_{a+b+c+d}^{a+b}} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!(a+b+c+d)!}$$

care este identică cu (2.8.4)* obținută prin altă metodă.

§ 12. Repartiția Poisson

Repartiția Poisson este un caz limită al repartiției binomiale. Pentru repartiția binomială

$$M(x) = n\theta = \mu.$$

Presupunem că n tinde către infinit și θ tinde la zero astfel încît $n\theta = \mu$ este o cantitate nenulă și finită. Avem

$$C_n^x \theta^x (1-\theta)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} =$$

$$= \frac{n(n-1)\dots(n-x+1)}{n^x} \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^{n-x} =$$

$$= \frac{n(n-1)\dots(n-x+1)}{n^x} \left(1 - \frac{\mu}{n}\right)^{-x} \frac{\mu^x}{x!} \left(1 - \frac{\mu}{n}\right)^{n}.$$

Pentru $n \to \infty$, obținem

$$(12.1) \qquad \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, \dots$$

Repartiția avînd funcția de frecvență dată de (12.1) poartă numele de *repartiție Poisson.*

---

* (2.8.4) are următoarea semnificație: capitolul 2, § 8, formula 4.

Scanned with OKEN Scanner

Avem

$$M(x) = \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu}\mu \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} = \mu,$$

$$M(x^2) = \sum_{x=0}^{\infty} x^2 \frac{\mu^x}{x!} e^{-\mu} = \mu e^{-\mu} \sum_{x=1}^{\infty} x \frac{\mu^{x-1}}{(x-1)!} =$$

$$= \mu e^{-\mu} \sum_{x=1}^{\infty} [(x-1)+1] \frac{\mu^{x-1}}{(x-1)!} =$$

$$= \mu e^{-\mu} \left[\mu \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\mu^{x-1}}{(x-1)!} \right] = \mu^2 + \mu$$

de unde

$$D^2(x) = \mu^2 + \mu - \mu^2 = \mu.$$

Observație. Faptul că $M(x) = D^2(x)$, este o caracteristică a repartiției Poisson.

§ 13. Teste privind repartiția Poisson

Dacă x este o variabilă binomială, atunci

$$P(x \leqslant k) = 1 - P\left(F < \frac{n-k}{k+1} \cdot \frac{\theta}{1-\theta}\right),$$

unde F este repartizată Fisher cu $2(k+1)$, respectiv $2(n-k)$ grade de libertate.

Dacă ținem seama că repartiția Poisson se obține din repartiția binomială cu $\theta = \frac{\mu}{n}$ și $n \to \infty$, pentru variabila Poisson obținem

$$P(x \leqslant k) = 1 - P\left(F < \frac{n-k}{n-\mu} \cdot \frac{\mu}{k+1}\right) = 1 - P\left(F < \frac{\mu}{k+1}\right),$$

unde F are $2(k+1)$ respectiv ∞ grade de libertate.
[Avem $F(f, \infty) = \chi^2(f)/f$]. Astfel

$$(13.1) \qquad P(x \leqslant k) = 1 - P\left(\frac{\chi^2}{2(k+1)} < \frac{\mu}{k+1}\right) = 1 - P(\chi^2 < 2\mu),$$

unde χ^2 are $2(k+1)$ grade de libertate.

Scanned with OKEN Scanner

(13.1) are o importanţă practică redusă. Cu toate acestea, ţinînd seama că (2.2.7) este o aproximaţie bună pentru funcţia de repartiţie a variabilei binomiale pentru $n\theta(1-\theta) > 9$, care în cazul Poisson corespunde lui $\mu > 9$, obţinem

$$P(x \leqslant k) \simeq \Phi\left(\frac{x+\frac{1}{2}-\mu}{\sqrt{\mu}}\right). \tag{13.2}$$

Se mai poate obţine o aproximaţie. În (2.4.6) am văzut că dacă dispersia variabilei este egală cu valoarea medie a variabilei, atunci rădăcina pătrată a variabilei are dispersia 1/4. Deci

$$P(x \leqslant k) \simeq \Phi\left(\frac{\sqrt{x+\frac{1}{2}}-\sqrt{\mu}}{1/4}\right) \simeq \Phi\left[2\left(\sqrt{x+\frac{1}{2}}-\sqrt{\mu}\right)\right]. \tag{13.3}$$

§ 14. Stabilitatea repartiţiei Poisson

Fie $x_1, x_2, \ldots x_n$ variabilele aleatoare independente (v.a.i) repartizate Poisson de parametri $\mu_1, \mu_2, \ldots$, respectiv μ_k. În aceste condiţii $x = \sum_1^k x_i$ este repartizată Poisson de parametru $\sum_1^k \mu_i$.

Demonstraţia acestor afirmaţii se face cu ajutorul funcţiei caracteristice.

§ 15. Compararea a două repartiţii Poisson

Într-o populaţie un eveniment s-a observat de x_1 ori, iar în altă populaţie, acelaşi eveniment, de x_2 ori. Ne propunem să verificăm ipoteza nulă conform căreia valoarea medie a numărului de realizări a evenimentului în cele două populaţii este aceeaşi, adică

$$H_0 : \lambda_1 = \lambda_2 = \lambda.$$

În ipoteza nulă x_1 şi x_2 sînt variabile aleatoare independente repartizate identic normal de medie λ şi dispersie λ. Prin urmare

Scanned with OKEN Scanner

$x_1 - x_2$ este repartizată normal de medie 0 și dispersie 2λ. Putem estima 2λ prin $(x_1 + x_2)$. Astfel în ipoteza nulă

$$\frac{x_1 - x_2}{\sqrt{x_1 + x_2}}$$

este repartizată $N(0, 1)$. Îmbunătățim această aproximație folosind corecția de continuitate analogă cu aceea folosită în problema celor două selecții. (Cazul bernoullian.) Dacă considerăm ipoteza alternativă $H_1 : \lambda_1 > \lambda_2$, vom respinge ipoteza nulă dacă

$$\frac{\left(x_1 - \frac{1}{2}\right) - \left(x_2 + \frac{1}{2}\right)}{\sqrt{x_1 + x_2}} > u_{1-\alpha}. \tag{15.1}$$

Un test exact poate fi construit urmărind construcția testului exact al lui Fisher pentru compararea a două populații binomiale.

Dacă este adevărată ipoteza nulă, atunci

$$P(x_1 = a, \ x_2 = b \ \text{și} \ x_1 + x_2 = a + b) = P(x_1 = a, \ x_2 = b) =$$

$$= P(x_1 = a)\, P(x_2 = b) = e^{-\lambda_1} \frac{\lambda_1^a}{a!} e^{-\lambda_2} \frac{\lambda_2^b}{b!} = e^{-2\lambda} \frac{\lambda^{a+b}}{a!b!}.$$

De asemenea

$$P(x_1 + x_2 = a + b) = e^{-2\lambda} \frac{(2\lambda)^{a+b}}{(a + b)!}$$

ceea ce ne permite să scriem

$$P(x_1 = a, \ x_2 = b \mid x_1 + x_2 = a + b) =$$

$$= \frac{P(x_1 = a, \ x_2 = b \ \text{și} \ x_1 + x_2 = a + b)}{P(x_1 + x_2 = a + b)} = \frac{\lambda^{a+b}}{(2\lambda)^{a+b}} \cdot \frac{(a + b)!}{a!b!} =$$

$$= C_{a+b}^{a} \left(\frac{1}{2}\right)^a = C_{a+b}^{a} \left(\frac{1}{2}\right)^a \left(1 - \frac{1}{2}\right)^{(a+b)-a}. \tag{15.2}$$

Ca și în testul exact al lui Fisher dorim să cunoaștem probabilitățile valorilor lui x_1 mai îndepărtate decît o valoare observată a. Astfel dacă $a > b$, ne interesează probabilitățile date de (15.2) pentru $x_1 = a + 1, \ a + 2, \ldots, a + b$, ultima fiind cea mai

Scanned with OKEN Scanner

mare valoare posibilă a lui x_1 deoarece corespunde valorii $x_2 = 0$. Probabilitatea căutată, suma probabilităților din coada repartiției, este

$$P(x_1 \geqslant a \mid a+b) = \sum_{x_1=a}^{a+b} C_{a+b}^{x_1} \left(\frac{1}{2}\right)^{x_1} \left(1-\frac{1}{2}\right)^{(a+b)-x_1} =$$

$$= 1 - \sum_{x_1=0}^{a-1} C_{a+b}^{x_1} \left(\frac{1}{2}\right)^{x_1} \left(1-\frac{1}{2}\right)^{(a+b)-x_1}.$$

Ținînd seama de proprietatea repartiției binomiale și de faptul că $\theta = \frac{1}{2}$, $\theta/(1-\theta) = 1$, obținem

$$P(x_1 \geqslant a \mid a+b) = P\left(F < \frac{b+1}{a}\right), \tag{15.3}$$

F fiind repartizată F cu $2a$, respectiv $2(b+1)$ grade de libertate.

Pentru testul unilateral cu ipoteza alternativă $\lambda_1 > \lambda_2$, P-cuantila este definită de

$$P\left(F(2a,\ 2(b+1)) < \frac{b+1}{a}\right) = P,$$

și cum $P(F < F_P) = P$, P este determinat de

$$F_P(2a,\ 2(b+1)) = \frac{b+1}{a}$$

sau ținînd seama de relația dintre cuantilele repartiției F

$$F_{1-P}(2(b+1),\ 2a)) = \frac{a}{b+1}.$$

§ 16. Compararea a două observații ce provin din populații caracterizate de variabile repartizate Poisson cu parametrii într-un anumit raport

Presupunem că în t_1 ore mașina A se defectează de x_1 ori, iar mașina B în t_2 ore de x_2 ori. Prezintă importanță verificarea ipotezei conform căreia probabilitatea numărului de defecțiuni pe unitate de timp este aceiași pentru fiecare mașină. Această ipoteză este echivalentă cu presupunerea că x_1 și x_2 sînt obser-

Scanned with OKEN Scanner

vaţii din populaţii Poisson ai căror parametrii λ_1 şi λ_2 se găsesc într-un anumit raport: $\frac{\lambda_1}{\lambda_2} = \frac{t_1}{t_2}$.

Testul exact din paragraful anterior pentru verificarea ipotezei $\lambda_1 = \lambda_2$ se extinde şi în acest caz.

Dacă ipoteza nulă este adevărată, probabilitatea ca $x_1 = a$, $x_2 = b$ este

$$P(x_1 = a,\ x_2 = b) = P(x_1 = a)\, P(x_2 = b) =$$

$$= e^{-\lambda_1} \frac{\lambda_1^a}{a!}\, e^{-\lambda_2} \frac{\lambda_2^b}{b!} = e^{-(\lambda_1+\lambda_2)} \frac{\lambda_1^a \lambda_2^b}{a!b!}.$$

Variabila aleatoare $x_1 + x_2$ este de asemenea repartizată Poisson de parametru $\lambda_1 + \lambda_2$ şi

$$P(x_1 + x_2 = a + b) = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^{a+b}}{(a+b)!}. \tag{16.2}$$

Ţinînd seama de (16.1) şi (16.2) obţinem

$$P(x_1 = a,\ x_2 = b \mid x_1 + x_2 = a + b) = \frac{P(x_1 = a,\ x_2 = b)}{P(x_1 + x_2 = a + b)} =$$

$$= \frac{(a+b)!}{a!b!} \cdot \frac{\lambda_1^a \lambda_2^b}{(\lambda_1 + \lambda_2)^{a+b}} = C_{a+b}^a \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^a \left(1 - \frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^b$$

care reprezintă probabilitatea ca într-un experiment binomial cu probabilitatea succesului $\lambda_1/(\lambda_1 + \lambda_2)$ în $a + b$ experienţe să avem a succese.

Presupunem că ne situăm în cazul în care valoarea observată a lui x_1 depăşeşte în ipoteza nulă valoarea sa medie şi invers valoarea observată a lui x_2 este inferioară valorii sale medii, adică $x_1/t_1 > x_2/t_2$.

Prin urmare trebuie să facem suma seriei binomiale pentru $x_1 = a,\ a + 1, \ldots, a + b$

$$P(x_1 \geqslant a \mid a + b) = \sum_{x_1=a}^{a+b} C_{a+b}^{x_1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{x_1} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{(a+b)-x_1}. \tag{16.4}$$

Coada superioară a repartiţiei binomiale este dată de

$$P(x \geqslant x_1) = P\left(F < \frac{n - x_1 + 1}{x_1} \cdot \frac{\theta}{1 - \theta}\right), \tag{16.5}$$

unde F are $2x_1$, respectiv $2(n - x_1 + 1)$ grade de libertate.

Scanned with OKEN Scanner

În (16.4) probabilitatea succesului este $\theta = \lambda_1/(\lambda_1 + \lambda_2)$ astfel încît $\theta/(1-\theta) = \lambda_1/\lambda_2$, iar (16.5) se scrie

$$P(x_1 \geqslant a \mid a+b) = P\left(F < \frac{a+b-a+1}{a} \cdot \frac{\lambda_1}{\lambda_2}\right),$$

unde F are $2a$, respectiv $2(b+1)$ grade de libertate.

Presupunem că această probabilitate este P_1. Atunci

$$\frac{b+1}{a} \cdot \frac{\lambda_1}{\lambda_2} = F_{P_1}(2a,\ 2(b+1)),$$

$$\frac{a}{b+1} \cdot \frac{\lambda_2}{\lambda_1} = F_{1-P_1}(2(b+1),\ 2a),$$

P_1 fiind determinat din tabelele cu repartiția F ca soluție a acestei ecuații pentru valorile observate a și b și valoarea λ_1/λ_2.

Scanned with OKEN Scanner

CAPITOLUL

3

TESTUL X^2

§ 1. Repartiția χ^2

Dacă $x_1, x_2, \ldots, x_n$ sînt variabile aleatoare independente repartizate normal $N(0,1)$ atunci spunem că variabila aleatoare $x = x_1^2 + x_2^2 + \ldots + x_n^2$ urmează o repartiție χ^2 cu n grade de libertate. Această variabilă are densitatea de repartiție

$$\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, \quad 0 \leqslant x < \infty. \tag{1.1}$$

Avem

$$M(x) = \sum_1^n M(x_i^2) = n \cdot 1 = n \quad \text{sau} \quad M(\chi_n^2) = n,$$

$$D^2(x) = \sum_1^n D^2(x_i^2) = n \cdot 2 = 2n \quad \text{sau} \quad D^2(\chi_n^2) = 2n$$

(1.1) pentru diferite valori ale lui n este reprezentată în figura 3.1.

Se observă că pentru $n > 2$ se accentuează apropierea de repartiția normală. Teorema limită centrală ne spune că variabila aleatoare x este asimptotic normală pentru valori mari ale lui n. Astfel putem obține cuantila de ordinul α al acestei repartiții din relația

$$\frac{x_\alpha - n}{\sqrt{2n}} \simeq u_\alpha$$

de unde

$$x_\alpha \simeq n + \sqrt{2n}\, u_\alpha \tag{1.2}$$

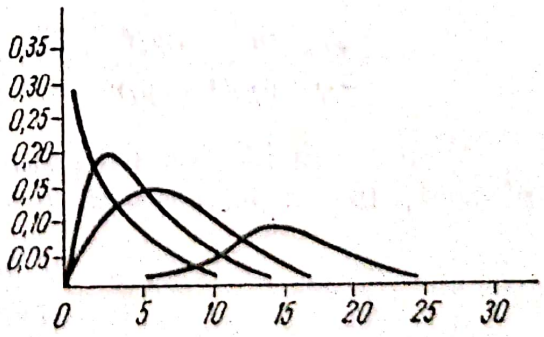


Fig. 3.1

O aproximație mai bună este

$$x_\alpha = \frac{1}{2}\,(\sqrt{2n-1} + u_\alpha)^2. \tag{1.3}$$

§ 2. Repartiția multinomială

Repartiția multinomială este o generalizare a repartiției binomiale în care în loc de două evenimente posibile A și A^c care pot rezulta în urma unei experiențe asupra unei variabile aleatoare X se consideră k evenimente $E_1, E_2, \ldots, E_k$, astfel încît

$$E_i \bigcap_{i \neq j} E_j = \emptyset, \quad \bigcup_1^k E_i = S, \quad \sum_1^k P(E_i) = 1.$$

Se arată că probabilitatea ca în n probe efectuate asupra variabilei aleatoare X, evenimentul $(E_i)_{1 \leqslant i \leqslant k}$ să se realizeze de n_i ori este

$$P_n(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\, n_2! \ldots n_k!}\, p_1^{n_1} p_2^{n_2}, \ldots p_k^{n_k}. \tag{2.1}$$

(2.1) reprezintă funcția de frecvență a *repartiției multinomiale*.

De asemenea pentru n și toți $(n_i)_{1 \leqslant i \leqslant k}$ suficienți de mari

$$P_n(n_1, n_2, \ldots, n_k) \simeq (2\pi n)^{(1-k)/2} \prod_1^k p_i^{-\frac{1}{2}} \exp\left(-\frac{X^2}{2}\right) \tag{2.2}$$

unde

$$X^2 = \sum_1^k \frac{(n_i - np_i)^2}{np_i}. \tag{2.3}$$

(2.3) se mai poate scrie

$$X^2 = \sum_1^k \frac{(n_i - np_i)^2}{np_i(1-p_i)}\,(1-p_i) = \sum_1^k \left(\frac{n_i - np_i}{\sqrt{np_i(1-p_i)}}\right)^2 (1-p_i).$$

Dacă considerăm un anumit eveniment E_i și-l privim ca eveniment, într-o experiență bernoulliană, atunci

$$\frac{n_i - np_i}{\sqrt{np_i(1-p_i)}}$$

este repartizată $N(0,1)$.

Scanned with OKEN Scanner

Deci (2.3) este o sumă ponderată de pătrate de variabile $N(0,1)$, ponderile fiind $(1 - p_i)$. Aceste variabile nu sînt independente deoarece $\sum_1^k (n_i - np_i) = 0$.

Astfel X^2 definit de (2.3) nu este variabila x definită în (1.1). Se arată ([3]) că pentru $n \to \infty$, (2.3) este repartizată χ^2_{k-1}.

În cazul unui test de semnificaţie bazat pe statistica X^2 dată de (2.3) respingem ipoteza emisă dacă (2.3) este mai mare decît $\chi^2_{\alpha;k-1}$.

§ 3. Tabele de contingenţă

Considerăm o selecţie de volum n, ale cărei elemente sînt clasificate după două atribute conform cărora fiecare individ poate fi descris ca aparţinînd la una din categoriile

$$A_1, A_2, \ldots, A_l$$

în raport cu atributul A şi simultan la una din categoriile

$$B_1, B_2, \ldots, B_k$$

în raport cu atributul B.

Fie p_{ij} probabilitatea ca un obiect să aibă atributele A_i şi B_j şi fie n_{ij} numărul elementelor din selecţie aparţinînd la categoriile A_i, B_j. Avem situaţia din tabelul 3.1.

Tabelul 3.1.

A \ B	B_1	B_2	. . . .	B_j	. . .	B_k	
A_1	n_{11}	n_{12}	. . . .	n_{1j}	. . .	n_{1k}	
A_2	n_{21}	n_{22}	. . . .	n_{2j}	. . .	n_{2k}	
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	
A_i	n_{i1}	n_{i2}	. . . .	n_{ij}	. . .	n_{ik}	n_i
.	.	.		.		.	
.	.	.		.		.	
.	.	.		.		.	
A_l	n_{l1}	n_{l2}	. . . .	n_{lj}	. . .	n_{lk}	
	$n \cdot j$						

Scanned with OKEN Scanner

Dacă punem

$$\sum_{j=1}^{k} n_{ij} = n_{i\cdot}, \qquad \sum_{i=1}^{l} n_{ij} = n_{\cdot j}$$

atunci

$$\sum_{i=1}^{l} n_{i\cdot} = \sum_{j=1}^{k} n_{\cdot j} = n.$$

Dacă probabilităţile p_{ij} sînt cunoscute, concordanţa dintre repartiţia ipotetică şi repartiţia empirică poate fi verificată cu ajutorul statisticii

$$X^2 = \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{(n_{ij} - np_{ij})^2}{np_{ij}}, \tag{3.1}$$

care este repartizată χ^2 cu $kl - 1$ grade de libertate.

În practică probabilităţile p_{ij} sînt necunoscute şi le vom estima cu ajutorul datelor.

Prezintă importanţă practică verificarea ipotezei conform căreia liniile şi coloanele de clasificare sînt independente, adică

$$P(A_i, B_j) = P(A_i)\, P(B_j)$$

sau dacă introducem notaţiile

$$p_{i\cdot} = \sum_{j=1}^{k} p_{ij}, \qquad p_{\cdot j} = \sum_{i=1}^{l} p_{ij}, \tag{3.2}$$

$$p_{ij} = p_{i\cdot}\, p_{\cdot j}$$

Estimaţiile de maximă verosimilitate pentru $p_{i\cdot}$ şi $p_{\cdot j}$ sînt

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n}, \quad 1 \leqslant i \leqslant l; \quad \hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n}, \quad 1 \leqslant j \leqslant k \tag{3.3}$$

şi ţinînd seama de acestea şi de (3.2) şi (3.1) obţinem

$$X^2 = \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{(n_{ij} - n_{i\cdot} n_{\cdot j}/n)^2}{n_{i\cdot} n_{\cdot j}/n}. \tag{3.4}$$

În acest caz (3.4) nu mai are kl-1 grade de libertate, deoarece cu ajutorul observaţiilor am estimat un număr de parametri.

Scanned with OKEN Scanner

S-au estimat l parametrii $p_{\cdot i}$, însă ținînd seama de relația

$$\sum_{i=1}^{l} p_{\cdot i} = 1$$

numai $l-1$ dintre acești parametri sînt independenți. Analog, se consumă $k-1$ grade de libertate pentru estimarea parametrilor $p_{i\cdot}$. Prin urmare X^2 definit de (3.4) are

$$(kl-1)-(k-1)-(l-1)=(k-1)(l-1)$$

grade de libertate.

§ 4. Verificarea normalității

Dacă variabila aleatoare X reprezintă caracteristica de calitate a unui produs, ne propunem să verificăm ipoteza

$$H_0 : X \in N(\mu, \sigma^2) \text{ față de } H_1 : X \notin N(\mu, \sigma^2).$$

Pentru a verifica această ipoteză să presupunem că $x_1,\ldots,x_n$ reprezintă măsurătorile asupra lui X. Intervalul valorilor posibile ale lui X îl împărțim în k intervale disjuncte $I_1, I_2,\ldots, I_k$, intervalul de ordin i avînd formula $I_i = (x_{i-1}, x_i)$, $1 \leqslant i \leqslant k$.

(4.1) Fie $p_i = P\{X \in I_k \mid H_0\}$

adică probabilitatea ca variabila să aparțină intervalului I_k cînd ipoteza H_0 este adevărată. Dacă H_0 este adevărată, atunci (4.1) devine:

$$p_i = F_0(x_i) - F_0(x_{i-1})$$

unde F_0 este funcția de repartiție corespunzătoare repartiției $N(\mu, \sigma^2)$,

$$F_0(x) = P(X < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt.$$

Din modul în care am definit probabilitățile p_i, $1 \leqslant i \leqslant k$, obținem

$$p_1 + p_2 + \ldots + p_k = 1.$$

Scanned with OKEN Scanner

Între funcția de repartiție F_0 și funcția lui Laplace.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt,$$

ale cărei valori sînt date în tabela 1, există relația

$$F_0(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Să presupunem că în n probe independente s-au observat n_1 valori în intervalul I_1,..., n_k valori în intervalul I_k. Numărul mediu de valori ale variabilei X în intervalul I_i este np_i.

Decizia cu privire la ipoteza H_0 se ia astfel:

a) dacă $X^2 \leqslant \chi^2_{\alpha;k-1}$ atunci se acceptă H_0;

b) dacă $X^2 > \chi^2_{\alpha;k-1}$ se respinge H_0 (se acceptă H_1) (unde statistica X^2 este dată de (2.3)).

În practică se întîlnește de obicei următoarea situație: pe baza datelor ce le avem la dispoziție, aproximăm grafic repartiția variabilei studiate și estimăm parametrii ce o caracterizează.

În această situație, nu se pot cunoaște exact valorile p_i ci doar valori aproximative ce depind de datele de selecție.

Atunci abaterile $(n_i - np_i)$ sînt supuse unui număr de legături liniare ce depinde de numărul parametrilor estimați ce caracterizează variabila X.

Fie „l" numărul acestor parametri, atunci numărul de legături liniare ce se impun abaterilor este $l + 1$.

Așadar, variabila X^2 scrisă mai înainte urmează la limită o repartiție χ^2 cu $k - (l + 1)$ grade de libertate.

În cazul repartiției normale $l = 2$ (media și dispersia); și deci avem de a face cu o repartiție χ^2 cu $k - 3$ grade de libertate.

În mod practic, testul X^2 pentru verificarea normalității, se aplică astfel:

a) valorile variabilei X se împart în intervale; fie acestea

$$(-\infty, x_1], \quad (x_1, x_2], \ldots, (x_{k-1}, x_k], \quad (x_k, +\infty)$$

unde

$$x_1 < x_2 < \ldots < x_k.$$

Se formează un tabel în care se trec în prima coloană aceste intervale.

Scanned with OKEN Scanner

b) Frecvențele absolute ale intervalelor se trec în coloana a doua a tabelului.

c) În intervalul $(-\infty, x_1]$ se numără toate valorile variabilei X care iau valori mai mici sau egale cu x_1, iar în intervalul $(x_k, +\infty)$ toate valorile lui X mai mari ca x_k.

d) În coloana a treia se trec extremitățile x_i din dreapta intervalelor.

e) Se calculează abaterile valorilor x_i față de media de selecție $\bar{x}$ și se trec în coloana a 4-a.

f) Se calculează abaterile normate, adică $(x_i - \bar{x})/s$ unde

$$s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2.$$

g) Din tabela 1 obținem valorile funcției Laplace în punctele $(x - \bar{x})/s$ și se trec în coloana a 5-a.

h) Se calculează probabilitățile p_i și se trec în coloana a 7-a.

Să observăm aici că:

$$p_1 = P(X \in I_1) = P\{X \in (-\infty, x_1)\} = P\{-\infty < X \leqslant x_1\} =$$

$$= F_0(x_1) - F_0(-\infty) = F_0(x_1) = \Phi\left(\frac{x_1 - \bar{x}}{s}\right),$$

$$p_i = F_0(x_i) - F_0(x_{i-1}) = \Phi\left(\frac{x_i - \bar{x}}{s}\right) - \Phi\left(\frac{x_{i-1} - \bar{x}}{s}\right), \quad 1 < i < k,$$

$$p_k = F_0(+\infty) - F_0(x_k) = 1 - F_0(x_k) = 1 - \Phi\left(\frac{x_k - \bar{x}}{s}\right).$$

i) în coloana a 8-a se înregistrează produsul dintre n și valorile p_i.

j) în coloana a 9-a se calculează abaterile $n_i - np_i$.

k) în coloana a 10-a se calculează $(n_i - np_i)^2$.

l) în coloana a 11-a se înregistrează expresiile $(n_i - np_i)^2/np_i$.

Sumînd rezultatele acestei coloane se obține expresia X^2.

Exemplu (E. Ursianu și V. Gh. Vodă: *Control statistic de calitate pentru fire relon* (I), „Materiale Plastice" Nr. 12/1971).

Scanned with OKEN Scanner

Tabelul 3.2

Interval	n_i	x_i	$x_i - \bar{x}$	$\frac{x_i - \bar{x}}{s}$	$\Phi\left(\frac{x_i - \bar{x}}{s}\right)$	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
$(-\infty, 445]$	5	445	—18	—2,01	0,0222	0,0222	3,5520	1,448	2,096704	0,5900
(445,448]	7	448	—15	—1,67	0,0475	0,0252	4,0480	2,952	8,714304	2,1527
(448,451]	3	451	—12	—1,34	0,0901	0,0426	6,8160	—3,816	14,561	2,1360
(451,454]	9	454	— 9	—1,01	0,1562	0,0661	10,5760	—1,576	2,483	0,2340
(454,457]	9	457	— 6	—0,67	0,2514	0,0952	15,2320	—6,23	38,37824	2,5497
(457,460]	19	460	— 3	—0,33	0,3707	0,1193	19,0880	—0,08	0,0077	0,0005
(460,463]	24	463	0	0	0,5000	0,1293	20,6880	3,32	11,0224	0,5327
(463,466]	21	466	3	0,33	0,6293	0,1293	20,6880	0,32	0,1024	0,0049
(466,469]	19	469	6	0,67	0,7486	0,1193	19,0880	—0,08	0,0064	0,0003
(469,472]	15	472	9	1,01	0,8438	0,0952	15,2320	—0,22	0,0529	0,0034
(472,475]	16	475	12	1,34	0,9099	0,0661	10,5760	+5,424	29,4764	2,7881
(475,478]	8	478	15	1,67	0,9525	0,0426	6,8160	1,184	1,4019	0,2093
(478,481]	3	481	18	2,01	0,9778	0,0253	4,0480	—1,048	1,098304	0,2712
$(481, +\infty)$	2	∞	∞	∞	1	0,0222	3,5520	—1,552	2,4087	0,6626

Scanned with OKEN Scanner

Se studiază normalitatea datelor privind caracteristica „torsiune" a firelor relon. S-au considerat pentru această caracteristică $n = 160$ date pe baza cărora s-au calculat estimațiile mediei și abaterii standard: $\bar{x} = 463$ și $s = 8{,}94$.

Datele au fost grupate în 14 intervale de lungime 3. Calculele sînt sintetizate în tabelul 3.2; însumînd ultima coloană găsim $X^2 = 14{,}241408$. Din tabelul 3 pentru $k - (l + 1) = 11$ grade de libertate și un risc $\alpha = 0{,}05$ găsim $\chi^2_{0{,}05;11} = 19{,}7$.

Cum $X^2 = 14{,}241408 < \chi^2_{0{,}05;11} = 19{,}7$ rezultă că repartiția variabilei „torsiune" este normală.

Scanned with OKEN Scanner

CAPITOLUL **4**

TESTE PENTRU EGALITATEA DISPERSIILOR ŞI MEDIILOR

§ 1. Dispersia de selecţie

Se arată că

$$s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x}_n)^2$$

este repartizată ca $\sigma^2\chi^2_{n-1}/(n-1)$, unde χ^2_{n-1} este o variabilă aleatoare repartizată χ^2 cu $(n-1)$ grade de libertate.

Printr-un abuz de scriere în acest paragraf vom nota prin n numărul gradelor de libertate.

Pentru a obţine un test de semnificaţie al ipotezei nule conform căreia $M(s^2) = \sigma_0^2$, faţă de alternativa conform căreia $M(s^2) = \sigma_1^2 > \sigma_0^2$, observăm că deoarece ns^2/σ_0^2 este repartizat (în ipoteza nulă) ca χ^2_n,

$$P\left(\frac{ns^2}{\sigma_0^2} > \chi^2_{n,1-\alpha} \mid M(s^2) = \sigma_0^2\right) = P(\chi^2_n > \chi^2_{n,1-\alpha}) = \alpha$$

astfel încît respingem ipoteza nulă dacă

$$\frac{ns^2}{\sigma_0^2} > \chi^2_{n,1-\alpha}$$

şi deci regiunea critică este

$$s^2 > \frac{\sigma_0^2\chi^2_{n,1-\alpha}}{n}.$$

Puterea acestui test este

$$\pi(\sigma_1^2) = P\left(s^2 > \frac{\sigma_0^2\chi^2_{n,1-\alpha}}{n} \mid M(s^2) = \sigma_1^2\right) =$$

$$= P\left(\frac{ns^2}{\sigma_1^2} > \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{n,1-\alpha}\right) = P\left(\chi^2 > \frac{\sigma_0^2}{\sigma_1^2}\chi^2_{n,1-\alpha}\right). \quad (1.1)$$

Scanned with OKEN Scanner

În cazul în care avem de verificat ipoteza H_0: $M(s^2) = \sigma_0^2$ față de alternativa H_1: $M(s^2) = \sigma_1^2 < \sigma_0^2$, regiunea critică este

$$s^2 < \frac{\sigma_0^2 \chi^2_{n,\alpha}}{n},$$

iar puterea testului

$$\pi(\sigma_1^2) = P\left(s^2 < \frac{\sigma_0^2 \chi^2_{n,\alpha}}{n} \mid M(s^2) = \sigma_1^2\right) =$$

$$= P\left(\frac{ns^2}{\sigma_1^2} < \frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,\alpha}\right) = P\left(\chi^2 < \frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,\alpha}\right). \tag{1.2}$$

În cazul în care avem de verificat ipoteza $H_0 : (s^2) = \sigma_0^2$ față de alternativa $H_1 : M(s^2) = \sigma_1^2 \neq \sigma_0^2$, regiunea critică este dată de cele două cozi (inferioară și superioară) fiecare de arie $\alpha/2$.

Ecuațiile (1.1) și (1.2) pot fi folosite pentru determinarea numărului gradelor de libertate necesare pentru obținerea unui anumite puteri.

Astfel dacă egalăm (1.1) cu $1 - \beta$, obținem

$$P\left(\chi^2 > \frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,1-\alpha}\right) = 1 - \beta$$

și ținînd seama că $P(\chi^2 > \chi^2_{n,\beta}) = 1 - \beta$, urmează că

$$\frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,1-\alpha} = \chi^2_{n,\beta} \quad \text{sau} \quad \frac{\sigma_1^2}{\sigma_0^2} = \frac{\chi^2_{n,1-\alpha}}{\chi^2_{n,\beta}}. \tag{1.3}$$

(1.3), pentru valori specificate ale lui $\frac{\sigma_1^2}{\sigma_0^2}$, α și β, poate fi rezolvată în raport cu n, folosind tabelul 3.

Pentru valori ale lui n mai mari decît 100, folosim aproximația (3.1.3.).

$$\chi^2_{n,P} = \frac{1}{2}(\sqrt{2n-1} + u_p)^2. \tag{1.4}$$

Ținînd seama de (1.4) în (1.3), obținem

$$\frac{\sigma_1^2}{\sigma_0^2} \simeq \frac{\frac{1}{2}(\sqrt{2n-1} + u_{1-\alpha})^2}{\frac{1}{2}(\sqrt{2n-1} + u_\beta)^2} \tag{1.5}$$

Scanned with OKEN Scanner

care rezolvată în raport cu n dă

$$n \simeq \frac{1}{2} + \frac{1}{2}\left(\frac{u_{1-\alpha} + u_{1-\beta}\sqrt{(\sigma_1^2/\sigma_0^2)}}{\sqrt{(\sigma_1^2/\sigma_0^2)} - 1} \right)^2 \tag{1.6}$$

§ 2. Raportul dispersiilor

Să presupunem că s_1^2 și s_2^2 sînt estimații independente pentru σ^2 bazate pe n_1, respectiv n_2 grade de libertate. Fie

$$F = \frac{s_1^2}{s_2^2}. \tag{2.1}$$

Ținînd seama că s_i^2 este repartizat, ca

$$\sigma^2\chi_{n_1}^2/n_i,\ i = 1, 2, \quad (2.1) \text{ se scrie}$$

$$F = \frac{\sigma^2\chi_{n_1}^2/n_1}{\sigma^2\chi_{n_2}^2/n_2} = \frac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2}. \tag{2.2}$$

Funcția de repartiție a variabilei F este tabelată în tabelele 5a, 5b și 5c.

Deoarece $0 \leqslant \chi^2 < \infty$, urmează că $0 \leqslant F < \infty$. Se poate arăta că

$$M(F) = \frac{n_2}{n_2 - 2}$$

și că mediana este $n_2(n_1 - 2)/n_1(n_2 + 2)$.

Tabelele 5 din anexă dă numai coada superioară a repartiției F. Pentru a obține puncte în coada inferioară procedăm după cum urmează. Deoarece

$$P\left(\frac{s_1^2}{s_2^2} < F_{P;n_1,n_2}\right) = P$$

urmează că

$$P\left\{\frac{s_2^2}{s_1^2} > \frac{1}{F_{1-P;n_1,n_2}}\right\} = 1 - P$$

sau

$$P\left(\frac{s_2^2}{s_1^2} < \frac{1}{F_{1-P;n_1,n_2}}\right) = P.$$

Scanned with OKEN Scanner

Cum

$$P\left(\frac{s_2^2}{s_1^2} < F_{P;n_2,n_1}\right) = P,$$

din ultimele două egalități obținem

$$F_{P;n_2,n_1} = \frac{1}{F_{1-P;n_1,n_2}}. \tag{2.3}$$

Dacă $s_1^2(s_2^2)$ este o estimație pentru $\sigma_1^2(\sigma_2^2)$ bazată pe $n_1(n_2)$ grade de libertate, atunci s_1^2/s_2^2 va fi repartizat ca

$$\frac{\frac{\sigma_1^2\chi_{n_1}^2}{n_1}}{\frac{\sigma_2^2\chi_{n_2}^2}{n_2}} = \left(\frac{\sigma_1}{\sigma_2}\right)^2 F_{n_1,n_2}. \tag{2.4}$$

Pentru a verifica ipoteza $H_0 : \sigma_1^2 = \sigma_2^2$ față de ipoteza alternativă $\sigma_1^2 > \sigma_2^2$, (2.4) ne spune că sub ipoteza nulă s_1^2/s_2^2 este repartizată F_{n_1,n_2}. Urmează că

$$P\left(\frac{s_1^2}{s_2^2} > F_{1-\alpha,n_1,n_2} \mid M(s_1^2) = \sigma_1^2 = \sigma_2^2 = M(s_2^2)\right) =$$

$$= P(F_{n_1,n_2} > F_{1-\alpha,n_1,n_2}) = 1 - (1-\alpha) = \alpha. \tag{2.5}$$

Prin urmare vom respinge ipoteza H_0 la pragul de semnificație α dacă

$$\frac{s_1^2}{s_2^2} > F_{1-\alpha,n_1,n_2}. \tag{2.6}$$

(2.6) definește regiunea critică pentru acest test.

Puterea testului pentru o valoare specificată a raportului $\frac{\sigma_1^2}{\sigma_2^2} = k$, este

$$\pi(k) = P\left(\frac{s_1^2}{s_2^2} > F_{1-\alpha;n_1,n_2} \mid M(s_1^2) = \sigma_1^2 = k\sigma_2^2 = kM(s_2^2)\right) =$$

$$= P\left(\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} > \frac{\sigma_2^2}{\sigma_1^2} F_{1-\alpha;n_1,n_2}\right) \tag{2.7}$$

Scanned with OKEN Scanner

sau ţinînd seama de (2.4)

$$\left(\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \text{ repartizată ca } F_{n_1, n_2}\right)$$

(2.7) se scrie

(2.8) $$\pi(k) = P\left(F_{n_1, n_2} > \frac{1}{k} F_{1-\alpha;\, n_1, n_2}\right).$$

Pentru a determina numărul de observaţii necesar pentru obţinerea unei anumite puteri $(1 - \beta)$, egalăm (2.8) cu $1 - \beta$

(2.9) $$P\left(F_{n_1, n_2} > \frac{1}{k} F_{1-\alpha;\, n_1, n_2}\right) = 1 - \beta.$$

(2.9) se mai poate scrie

$$P\left(F_{n_1, n_2} < \frac{1}{k} F_{1-\alpha;\, n_1, n_2}\right) = \beta$$

de unde

$$\frac{1}{k} F_{1-\alpha;\, n_1, n_2} = F_{\beta;\, n_1, n_2}$$

sau

(2.10) $$k = \frac{F_{1-\alpha;\, n_1, n_2}}{F_{\beta;\, n_1, n_2}} = F_{1-\alpha;\, n_1, n_2} \cdot F_{1-\beta;\, n_2, n_1},$$

care poate fi rezolvată iterativ cu ajutorul tabelelor 5.

§ 3 a. Testul lui Bartlett pentru egalitatea mai multor dispersii

Testul lui Bartlett este o generalizare a celor discutate în paragraful anterior. Cu ajutorul acestui test se verifică ipoteza nulă

(3.1) $$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2.$$

Dacă este adevărată ipoteza H_0, atunci

$$s^2 = \sum_1^k n_i s_i^2 \Big/ \sum_1^k n_i$$

este repartizată ca $\sigma^2 \chi_n^2/n$, unde $\sum_1^k n_i = n$. Astfel repartiţia

Scanned with OKEN Scanner

oricărei funcţii de $\frac{s_1}{s^2}, \ldots, \frac{s_k^2}{s^2}$, depinde numai de parametrii n_1, $n_2, \ldots, n_k$ şi k. Bartlett a arătat că raportul

$$\frac{-\sum_1^k n_i \ln \frac{s_i^2}{s^2}}{1 + \frac{1}{k-1}\sum_1^k \left(\frac{1}{n_i} - \frac{1}{n}\right)} \tag{3.2}$$

este repartizat, în ipoteza nulă, χ^2 cu $(k-1)$ grade de libertate.

Exemplul 1 (E. Ursianu şi V. Gh. Vodă: *Control statistic de calitate pentru fire relon* I, „Materiale plastice" nr. 12, 1971).

Pentru alcătuirea unor planuri de control statistic de recepţie pentru fire tip relon s-a studiat stabilitatea procesului tehnologic relativ la caracteristicile de calitate ale produsului respectiv.

Aceasta revine la verificarea ipotezei statistice (3.1). Pentru caracteristica „alungire" a firului de relon s-au cules date pe o perioadă de 11 luni.

Tabelul 4.1 sintetizează numărul gradelor de libertate şi dispersiile de selecţie pentru $k = 11$ luni.

Tabelul 4.1

Nr.	ν_i	s_i^2	s_i
1	151	6,8306	2,61
2	159	7,3437	2,70
3	159	8,9359	2,98
4	159	7,4673	2,73
5	151	6,5592	2,56
6	159	7,9302	2,81
7	127	6,4025	2,53
8	103	8,5293	2,92
9	167	6,9790	2,64
10	119	6,1055	2,47
11	103	9,4061	3,06

Scanned with OKEN Scanner

Calculele furnizează rezultatele (Olivetti, Programma 101):

$$\nu = \sum_1^{11} \nu_i = 1557; \quad \sum_1^{11} \frac{1}{\nu_i} = 0{,}080081; \quad \nu \ln s^2 = 2{,}009193,$$

$$\sum_1^{11} \nu_i \ln s_i^2 = 3115{,}496553.$$

Statistica lui Bartlett (3.2) are valoarea $X^2 = 13{,}165788$. Cum $\chi^2_{0{,}05:10} = 18{,}3$ ipoteza egalității dispersiilor se acceptă. Așadar, se poate trage concluzia că procesul tehnologic este stabil pentru această caracteristică.

Verificarea egalității dispersiilor permite estimarea dispersiilor teoretice prin valoarea lor comună pe cînd inegalitatea lor creează complicații din punct de vedere practic.

În cazul analizat estimația abaterii standard teoretice este:

$$\hat{\sigma} = \sqrt{\frac{\nu_1 s_1^2 + \ldots + \nu_k s_k^2}{\nu_1 + \ldots + \nu_k}} = \sqrt{\frac{151 \cdot 6{,}8306 + \ldots + 103 \cdot 9{,}4061}{151 + \ldots + 130}} = 2{,}73.$$

În cazul că nu se dispune de calculator, este indicat să se procedeze ca în:

Exemplul 2 (D. V. Iliescu și V. Gh. Vodă: *Calitatea și problema stabilității în funcționarea utilajelor*, „Calitatea producției și metrologie" nr. 5, 1971, pp. 288—290).

Se studiază stabilitatea procesului tehnologic relativ la caracteristica „termostabilitate la 177°C" a rețelelor cord ce se folosesc la fabricarea anvelopelor.

În tabelul 4.2 sînt sintetizate datele necesare intrării în formula (3.2).

Tabelul 4.2

nr.	s_i^2	n_i	ν_i $(n_i - 1)$	$\nu_i s_i^2$	$\lg s_i^2$	$\nu_i \lg s_i^2$	$1/\nu_i$
1	3,842	96	95	364,990	0,584	55,480	1/95
2	5,552	72	71	312,062	0,742	52,682	1/71
3	7,075	80	79	558,925	0,848	67,071	1/79
4	10,368	72	71	736,128	1,016	72,136	1/71

Scanned with OKEN Scanner

Tabelul 4.2 (continuare)

nr.	s_i^2	n_i	ν_i (n_i-1)	$\nu_i s_i^2$	$\lg s_i^2$	$\nu_i \lg s_i^2$	$1/\nu_i$
5	6,368	72	71	450,850	0,803	57,013	1/71
6	6,350	96	95	603,250	0,803	76,285	1/95
7	4,709	96	95	447,355	0,673	63,935	1/95
8	5,664	88	87	492,768	0,753	65,511	1/86
9	4,285	72	71	304,235	0,631	44,801	1/71
10	4,410	80	79	348,390	0,644	50,876	1/71
11	5,118	72	71	369,058	0,716	50,836	1/71
			885	5068,011		1510,24	0,139

(se foloseşte formula aproximativă $\ln X \simeq 2{,}3 \lg X$)

În cazul nostru $\chi^2 \simeq 32$; cum $\chi^5_{0{,}05;\,10} = 18{,}3$ ipoteza omogenităţii dispersiilor se respinge.

§ 3 b. Testul lui Cochran pentru egalitatea mai multor dispersii

În cazul în care numărul de măsurători este acelaşi pentru toate cele k selecţii este indicat să se aplice un alt test mai rapid decît cel al lui Bartlett şi anume testul lui Cochran; el constă în următoarele:

— se calculează dispersiile de selecţie $s_1^2, s_2^2, \ldots, s_k^2$;

— se alege $\max\limits_{1 \leqslant i \leqslant k} s_i^2$;

— se formează raportul $Q = \dfrac{\max\limits_{1 \leqslant i \leqslant k} s_i^2}{\sum\limits_{i=1}^{k} s_i^2}$;

— din tabelele 9 se extrage $Q_{\alpha;\,k,\,n-1}$; dacă $Q > Q_{\alpha;\,k,\,n-1}$ se respinge ipoteza omogenităţii dispersiilor.

Exemplul 3. (Al. C. Dorin, A. Sandu şi V. Gh. Vodă: *Aprecierea stabilităţii în funcţionare a maşinilor — unelle cu comandă numerică*, „Calitatea producţiei şi metrologie“ Nr. 5/1972).

Scanned with OKEN Scanner

În tabelul 4.3 sînt prezentate erorile obținute (în microni) în urma măsurării deplasărilor după axele de coordonare OX și OY ale unei mașini de alezat și frezat cu comandă numerică; aceste măsurători au fost înregistrate pentru a se vedea dacă mașina-unealtă își menține precizia în timp.

Tabelul 4.3

Nr. grupe / Nr. curent	1	2	3	4	5	6	7
1	6	0	16	17	18	10	1
2	12	16	10	16	14	15	10
3	8	4	9	10	0	4	7
4	6	4	0	6	4	10	9
5	4	3	3	6	4	4	7
6	3	0	16	6	12	0	3
7	11	14	3	12	6	15	10
8	0	6	13	15	10	3	4
9	18	12	0	0	19	8	6
10	7	5	10	12	0	13	15
11	0	10	5	8	5	4	3
12	7	10	3	12	17	14	18
13	0	5	4	0	0	0	0
14	5	8	6	12	7	5	3

În acest caz $k = 7$, $n_i = 14$ $(i = 1, \ldots, 7)$, $\nu_i = n_i - 1 = 13$. Dispersiile obținute sînt:

i	1	2	3	4	5	6	7
s_i^2	25,5581	24,2179	29,2220	27,2885	37,9446	28,7221	26,8930

Se vede că $\max\limits_{1 \leqslant i \leqslant 7} s_i^2 = 37{,}9446$; $\sum\limits_{1}^{7} s_i^2 = 199{,}8462$, $n_1 = n_2 = \ldots = n_7 = n = 14$ și deci

$$Q = \frac{37{,}9446}{199{,}8462} = 0{,}1896$$

Scanned with OKEN Scanner

Din tabele se extrage:

$$Q_{0,05;7:13} \simeq 0,250.$$

Se trage deci concluzia omogeneității dispersiilor, deci mașina este precisă la momentul testării.

§ 4. Testul t

a. Fie $x_1, x_2, \ldots, x_n$ o selecție dintr-o populație $N(\mu, \sigma^2)$. Ne propunem să verificăm ipoteza nulă

$$H_0: \mu = \mu_0.$$

Putem scrie

$$\sum_1^n (x_i - \mu)^2 = n(\bar{x} - \mu)^2 + \sum_1^n (x_i - \bar{x}_n)^2$$

unde $n(\bar{x} - \mu)^2$ și $\sum_1^n (x_i - \bar{x}_n)^2$ sînt independente și repartizate ca $\sigma^2\chi_1^2$, respectiv $\sigma^2\chi_{n-1}^2$. Urmează că

$$M[n(\bar{x} - \mu)^2] = \sigma^2, \tag{4.1}$$

$$M\left[\sum_1^n (x_i - \bar{x}_n)^2\right] = (n-1)\,\sigma^2. \tag{4.2}$$

Punînd

$$s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2 \tag{4.3}$$

și

$$s_*^2 = n(\bar{x} - \mu)^2, \tag{4.4}$$

urmează că $s^2(s_*^2)$ este repartizată ca $\sigma^2\chi_{n-1}^2/(n-1)\,[\sigma^2\chi_1^2/1]$ și ținînd seama de (2.2)

$$s_*^2/s^2$$

este repartizată $F_{1,n-1}$.

Deoarece nu cunoaştem media μ, nu putem calcula pe s_*^2. Vom defini statistica

$$s_{**}^2 = n(\bar{x} - \mu_0)^2. \tag{4.5}$$

Avem

$$s_{**}^2 = n(\bar{x} - \mu_0)^2 = n\,[(\bar{x} - \mu) + (\mu - \mu_0)]^2 =$$

$$= n(\bar{x} - \mu)^2 + n(\mu - \mu_0)^2 + 2n(\mu - \mu_0)\,(\bar{x} - \mu)$$

şi ţinînd seama că

$$M\,[2n(\mu - \mu_0)\,(\bar{x} - \mu)] = 2n(\mu - \mu_0)\,M(\bar{x} - \mu) = 0,$$

şi de (4.1), obţinem

$$M(s_{**}^2) = \sigma^2 + n(\mu - \mu_0)^2. \tag{4.6}$$

În cazul în care este adevărată ipoteza H_0, $s_*^2 = s_{**}^2$ şi

$$\frac{s_{**}^2}{s^2} \text{ este repartizată } F_{1,n-1}. \tag{4.7}$$

Dacă ipoteza H_0 este falsă, numărătorul din (4.7) va avea o valoare medie mai mare decît σ^2.

Ipoteza H_0 mai poate fi verificată după cum urmează: Se ştie (vezi G. Ciucu, V. Craiu [3]) că media de selecţie $\bar{x}$ este independentă de dispersia de selecţie

$$s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x}_n)^2.$$

Dacă este adevărată ipoteza H_0, atunci $M(x) = \mu_0$ şi

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

este repartizat $N(0, 1)$.

Scanned with OKEN Scanner

De asemenea deoarece $\sum_1^n (x_i - \bar{x})^2$ este repartizat $\sigma^2\chi^2_{n-1}$, s^2 va fi repartizat $\sigma^2\chi^2_{n-1}/(n-1)$. Urmează că

$$\frac{(\bar{x} - \mu_0)/\sigma/\sqrt{n})}{s/\sigma} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{4.8}$$

este repartizat student cu $(n-1)$ grade de libertate.

Acest procedeu este acelaşi cu cel bazat pe s^2_{**}/s^2 dat de (4.7), pentru că pătratul lui (4.8) este

$$\frac{n(\bar{x} - \mu_0)^2}{s^2} = \frac{s^2_{**}}{s^2}$$

şi deoarece $t^2_{n-1} = F_{1,n-1}$. Într-adevăr dacă în (2.2) facem $n_1 = 1$, numărătorul devine $\chi^2_1/1$ sau u^2, unde u este repartizat $N(0, 1)$. Deci

$$F_{1,n_2} = \frac{u^2}{\chi^2_{n_2}/n_2},$$

care este pătratul unei variabile repartizate student cu n_2 grade de libertate.

Pentru testul de alternativa $\mu > \mu_0$, regiunea critică (1.3.6) este

$$\bar{x} > \mu_0 + t_{1-\alpha}\frac{s}{\sqrt{n}}, \tag{4.9}$$

pentru alternativa $\mu < \mu_0$,

$$\bar{x} < \mu_0 + t_{\alpha}\frac{s}{\sqrt{n}}, \tag{4.10}$$

iar pentru alternativa $\mu \neq \mu_0$ (1.5.5) dă regiunea critică

$$\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| > t_{1-\frac{\alpha}{2}}. \tag{4.11}$$

Aceste regiuni critice sînt analoage cu acelea obţinute cînd σ era presupus cunoscut. Singura diferenţă este că folosirea lui s în locul lui σ implică că am folosit repartiţia t în locul repartiţiei normale.

Scanned with OKEN Scanner

b. Fie $x_{11}, x_{12}, \ldots, x_{1n_1}(x_{21}, x_{22}, \ldots, x_{2n_2})$ o selecție dintr-o populație $N(\mu_1, \sigma^2)\,[N(\mu_2, \sigma^2)]$ cu σ^2 necunoscut. Ne propunem să verificăm ipoteza

$$H_0: \mu_1 = \mu_2.$$

Pentru fiecare selecție, avem:

$$\sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 = n_i(\bar{x}_{n_i} - \mu_i)^2 + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{n_i})^2.$$

În acest caz (4.1) și (4.2) devin

$$M\,[n_i(\bar{x}_{n_i} - \mu_i)^2] = \sigma^2, \tag{4.1'}$$

$$M\left[\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{n_i})^2\right] = (n_i - 1)\,\sigma^2 \qquad i = 1, 2 \tag{4.2'}$$

Statistica

$$s^2 = \frac{\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_{n_1})^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_{n_2})^2}{n_1 - 1 + n_2 - 1} \tag{4.12}$$

are valoarea medie egală cu σ^2 și este repartizată ca

$$\sigma^2\chi^2_{n_1-1+n_2-1}/(n_1 - 1 + n_2 - 1).$$

Deoarece $n_i(\bar{x}_{n_i} - \mu_i)^2$ este repartizată ca $\sigma^2\chi_1^2$, urmează că

$$n_1(\bar{x}_{n_1} - \mu_1)^2 + n_2(\bar{x}_{n_2} - \mu_2)^2 \tag{4.13}$$

este repartizată ca $\sigma^2\chi_2^2$.

(4.13) este independentă de (4.12).

Fie

$$\mu = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}, \tag{4.14}$$

$$\bar{x} = \frac{n_1\bar{x}_{n_1} + n_2\bar{x}_{n_2}}{n_1 + n_2}, \tag{4.15}$$

și

$$\delta_i = \mu_i - \mu. \tag{4.16}$$

Scanned with OKEN Scanner

Deoarece

$$\bar{x}_{n_i} - \mu_i = (\bar{x}_{n_i} - \bar{x}) - (\mu_i - \mu) + (\bar{x} - \mu) =$$
$$= [(\bar{x}_{n_i} - \bar{x}) - \delta_i] + (\bar{x} - \mu),$$

putem scrie

$$\sum_1^2 n_i(\bar{x}_{n_i} - \mu_i)^2 = \sum_1^2 n_i[(\bar{x}_{n_i} - \bar{x}) - \delta_i]^2 + (n_1 + n_2)(\bar{x} - \mu)^2 \tag{4.17}$$

deoarece

$$\sum_1^2 2n_i(\bar{x}_{n_i} - \bar{x} - \delta_i)(\bar{x} - \mu) = 2(\bar{x} - \mu)\sum_1^2 n_i(\bar{x}_{n_i} - \bar{x} - \delta_i) =$$
$$= 2(\bar{x} - \mu)[n_1\bar{x}_1 + n_2\bar{x}_2 - (n_1 + n_2)\bar{x} - (n_1\delta_1 + n_2\delta_2) = 0$$

și s-a ținut seama de (4.14) — (4.16).

(4.13) ne spune că membrul stîng al relației (4.17) este repartizat ca $\sigma^2\chi_2^2$. În membrul drept avem o sumă de pătrate și o variabilă la pătrat. Suma de pătrate conține două variabile $\bar{x}_{n_1}$ și $\bar{x}_{n_2}$ supuse la condiția (4.15) și deci cu $2 - 1 = 1$ grade de libertate. Variabila la pătrat este $\bar{x}$ și deci avem un singur grad de libertate. Suma de pătrate și pătratul variabilei sînt independente și repartizate ca $\sigma^2\chi_1^2$.

Fie

$$s'^2 = \sum_1^2 n_i[(\bar{x}_{n_i} - \bar{x}) - \delta_i]^2. \tag{4.18}$$

Avem

$$M(s'^2) = M(\sigma^2\chi_1^2) = \sigma^2$$

astfel încît s'^2 și s^2 sînt estimații pentru σ^2, iar

$$\frac{s'^2}{s^2} \sim F_{1,\, n_1+n_2-2}.$$

Nu putem calcula s'^2 deoarece în expresia sa intervine δ_i care sînt necunoscute. Definim

$$s_*'^2 = \sum_1^2 n_i(\bar{x}_{n_i} - \bar{x})^2. \tag{4.19}$$

Scanned with OKEN Scanner

Deoarece

$$\sigma^2 = M(s'^2) = M[n_1(\bar{x}_{n_1} - \bar{x})^2 + n_2(\bar{x}_{n_2} - \bar{x})^2 + n_1\delta_1^2 + n_2\delta_2^2 -$$
$$- 2n_1\delta_1(\bar{x}_{n_1} - \bar{x}) - 2n_2\delta_2(\bar{x}_{n_2} - \bar{x})] =$$
$$= M\left[\sum_1^2 n_i(\bar{x}_{n_i} - \bar{x})^2\right] + n_1\delta_1^2 + n_2\delta_2^2 -$$
$$- 2n_1\delta_1 M(\bar{x}_{n_1} - \bar{x}) - 2n_2\delta_2 M(\bar{x}_{n_2} - \bar{x}),$$

obținem

$$M\left[\sum_1^2 n_i(\bar{x}_{n_i} - \bar{x})^2\right] = M(s_*'^2) = \sigma^2 + (n_1\delta_1^2 + n_2\delta_2^2) =$$
$$= \sigma^2 + \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)^2. \tag{4.20}$$

Prin urmare, dacă ipoteza H_0: $\mu_1 = \mu_2$ este adevărată, atunci $s_*'^2 = s'^2$, $s_*'^2 \sim \sigma^2\chi_1^2$ și

$$\frac{s_*'^2}{s^2} \sim F_{1,\, n_1+n_2-2}. \tag{4.21}$$

Ipoteza H_0 mai poate fi verificată după cum urmează. Mediile de selecție $\bar{x}_{n_1}$ și $\bar{x}_{n_2}$ sînt independente de $\sum_1^{n_1}(x_{1i} - \bar{x}_{n_1})^2$ și $\sum_1^{n_2}(x_{2j} - \bar{x}_{n_2})^2$ și prin urmare și de s^2 definit în (4.12). s^2 este repartizat ca $\sigma^2\chi^2_{n_1+n_2-2}/(n_1 + n_2 - 2)$. Dacă este adevărată ipoteza H_0: $\mu_1 = \mu_2$, atunci $\bar{x}_{n_1} - \bar{x}_{n_2}$ va fi repartizat normal de medie zero și dispersie

$$\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Urmează că

$$\frac{\bar{x}_{n_1} - \bar{x}_{n_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Scanned with OKEN Scanner

și deci

$$(4.22)\qquad \frac{\bar{x}_{n_1}-\bar{x}_{n_2}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\cdot\frac{1}{s/\sigma}=\frac{\bar{x}_{n_1}-\bar{x}_{n_2}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}.$$

Statistica definită în (4.22) este identică cu cea definită în (4.21).

Într-adevăr

$$s'^2=n_1(\bar{x}_{n_1}-\bar{x})^2+n_2(\bar{x}_{n_2}-\bar{x})^2=$$

$$=n_1\left(\bar{x}_{n_1}-\frac{n_1\bar{x}_{n_1}+n_2\bar{x}_{n_2}}{n_1+n_2}\right)^2+n_2\left(\bar{x}_{n_2}-\frac{n_1\bar{x}_{n_1}+n_2\bar{x}_{n_2}}{n_1+n_2}\right)^2=$$

$$=\left(\frac{\bar{x}_{n_1}-\bar{x}_{n_2}}{n_1+n_2}\right)^2(n_1n_2^2+n_1^2n_2)=\frac{(\bar{x}_{n_1}-\bar{x}_{n_2})^2}{\frac{1}{n_1}+\frac{1}{n_2}},$$

astfel încît

$$\frac{s'^2}{s^2}=\frac{\bar{x}_{n_1}-\bar{x}_{n_2}}{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},$$

care este pătratul lui (4.22). Deoarece

$$t^2_{n_1+n_2-2}=F_{1,n_1+n_2-2},$$

cele două teste sînt identice.

c. Fie $x_{11}, x_{12}, \ldots, x_{1n_1}(x_{21}, x_{22}, \ldots, x_{2n_2})$ o selecție dintr-o populație $N(\mu_1, \sigma_1^2)$ $[N(\mu_2, \sigma_2^2)]$. Ne propunem să verificăm ipoteza $\mu_1=\mu_2$, dispersiile σ_1^2 și σ_2^2 fiind necunoscute.

$\bar{x}_{n_1}=\frac{1}{n_1}\sum_1^{n_1}x_{1i}$, $\bar{x}_{n_2}=\frac{1}{n_2}\sum_1^{n_2}x_{2i}$ sînt repartizate normal de dispersii $\frac{\sigma_1^2}{n_1}$, $\frac{\sigma_2^2}{n_2}$, astfel încît

$$(4.23)\qquad \frac{\bar{x}_{n_1}-\bar{x}_{n_2}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Scanned with OKEN Scanner

este repartizată $N(0, 1)$. Deoarece nu cunoaștem dispersiile σ_1^2 și σ_2^2, trebuie să determinăm repartiția statisticii (4.23) cînd înlocuim σ_i^2 prin $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{n_i})^2$, adică a statisticii

$$\frac{(\bar{x}_{n_1} - \bar{x}_{n_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \tag{4.24}$$

Mai întîi ne vom ocupa de o problemă mai generală. Fie

$$S^2 = a_1 s_1^2 + a_2 s_2^2 + \ldots + a_k s_k^2, \tag{4.25}$$

unde $s_1^2, \ldots, s_k^2$ sînt independente, iar a_i constante cunoscute.

Deoarece

$$\frac{(n_i - 1)\, s_i^2}{\sigma_i^2} \sim \chi^2(n_i - 1), \quad 1 \leqslant i \leqslant k$$

urmează că

$$D^2(s_i^2) = D^2\left(\frac{\sigma_i^2 \chi^2_{(n_i-1)}}{n_i - 1}\right) = (\sigma_i^2)^2\, D^2 = \left(\frac{\chi^2_{n_i-1}}{n_i - 1}\right) = 2\,\frac{(\sigma_i^2)^2}{n_i - 1}. \tag{4.26}$$

Deci

$$D^2(S^2) = D^2\left(\sum_1^k a_i s_i^2\right) = \sum_1^k a_i^2 D^2(s_i) = \sum_1^k a_i^2 \frac{2(\sigma_i^2)^2}{n_i - 1}. \tag{4.27}$$

Vom aproxima repartiția lui S^2 definită în (4.25), printr-o repartiție de pătrate de medii cu același număr de grade de libertate, n^*. Vom face media și dispersia repartiției ce aproximează, egală cu media și dispersia repartiției inițiale.

Dispersia repartiției de pătrate cu aceeași medie ca $S^2\left(\sum_1^k a_i \sigma_i^2\right)$ și grade de libertate n^*, ținînd seama de (4.26) este

$$\frac{2\left(\sum_1^k a_i \sigma_i^2\right)^2}{n^*}. \tag{4.28}$$

Scanned with OKEN Scanner

Egalînd (4.27) cu (4.28) și rezolvînd în raport cu n^*, obținem

$$n^* = \frac{\left(\sum_1^k a_i \sigma_i^2\right)^2}{\sum_1^k a_i^2 \frac{(\sigma_i^2)^2}{n_i - 1}}. \tag{4.29}$$

Deoarece în practică nu sînt cunoscute dispersiile σ_i^2 vom folosi dispersiile de selecție s_i^2. În acest mod n^* capătă forma

$$n^* \simeq \frac{\left(\sum_1^k a_i s_i^2\right)^2}{\sum_1^k a_i^2 \frac{(s_i^2)^2}{n_i - 1}}. \tag{4.30}$$

În cazul particular, al statisticii (4.24),

$$a_1 = \frac{1}{n_1}, \quad a_2 = \frac{1}{n_2}$$

de unde

$$n^* \simeq \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1}\right)^2 \frac{(s_1^2)^2}{n_1 - 1} + \left(\frac{1}{n_2}\right)^2 \frac{(s_2^2)^2}{n_2 - 1}}. \tag{4.31}$$

Punînd

$$c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{4.32}$$

(4.31) se poate pune sub forma

$$\frac{1}{n^*} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}. \tag{4.33}$$

Derivînd (4.33) în raport cu c, iar apoi egalînd cu zero, obținem că cea mai mare valoare pe care o poate lua n^* este $(n_1 - 1) + (n_2 - 1)$ și aceasta se întîmplă cînd

$$\frac{s_1^2}{n_1(n_1 - 1)} = \frac{s_2^2}{n_2(n_2 - 1)}.$$

Scanned with OKEN Scanner

Cea mai mare valoare pe care poate s-o ia n^* este

$$\min [n_1 - 1, n_2 - 1]$$

și aceasta se întîmplă cînd $1 - c$ sau c se apropie de zero, sau același lucru cînd

$$\frac{s_2^2}{n_2} \Big/ \frac{s_1^2}{n_1} \quad \text{sau} \quad \frac{s_1^2}{n_1} \Big/ \frac{s_2^2}{n_2}$$

se apropie de zero.

§ 5. Testul lui Dixon

În cazul în care dispunem de m selecții de același volum asupra unei anumite caracteristici pentru eliminarea valorilor care diferă semnificativ de celelalte se folosește *testul lui Dixon*, care constă în următoarele:

1) pentru fiecare selecție se determină amplitudinea (R_j), adică se face diferența între valoarea cea mai mare și valoarea cea mai mică apărută în eșantionul respectiv.

2) se calculează statistica

$$Q = \frac{R_{max}}{\sum_1^m R_j}.$$

3) se compară valoarea calculată Q cu coeficientul Q_t care se găsește în tabele în funcție de volumul eșantionului și de riscul ales (în general 0,05).

Decizia de eliminare a valorilor extreme se ia dacă

$$Q > Q_t.$$

În caz contrar ($Q \leqslant Q_t$) nu se face nici o eliminare.

Aplicații ale testelor F și t

Cele ce urmează sînt extrase din lucrarea „Folosirea unor metode statistico-matematice în studierea rezultatelor a două metode de analiză în controlul rețelelor cord (I—II) de V. Gh.

Scanned with OKEN Scanner

Vodă, D. V. Iliescu, S. Comănici și Ligia Cojocaru, „Materiale Plastice" nr. 11/1971.

Rețelele cord poliamidice sînt utilizate ca inserție la fabricarea anvelopelor. Rețeaua este o țesătură din fire relon și fire bumbac.

Elementul de rezistență al rețelei îl constituie firele de relon cord dispuse ca urzeală. Firele de bumbac au un rol secundar: menținerea paralelismului firelor de relon pentru a forma o suprafață plană.

Firele de relon din care sînt executate rețelele trebuie să îndeplinească anumite condiții tehnice, cum ar fi: sarcina de rupere, alungirea la rupere, contracția, termostabilitatea etc.

Aceste condiții tehnice sînt stabilite în STAS-uri și norme interne, în funcție de sortimentul fabricat. Tot prin aceste normative sînt reglementate și condițiile de recepție, precum și modul de efectuare a determinărilor privind caracteristicile tehnice de calitate.

Determinările de laborator sînt executate pe fire extrase din rețea. În acest scop la două role din lot se țese în plus un metru de probă. Acest metru de probă se împarte în două. O jumătate de metru de probă servește la determinările de laborator ale furnizorului, iar cealaltă jumătate la determinările de laborator ale beneficiarului pentru recepționarea produsului.

Furnizorul, pe baza informațiilor date de jumătatea de probă, încadrează lotul într-o anumită clasă de calitate.

Beneficiarul, pe baza informațiilor date de jumătatea sa de probă, acceptă sau nu lotul la calitatea livrată.

Decizia, atît la furnizor, cît și la beneficiar, se ia pe baza valorii medii a unui număr „*n*" de determinări efectuate prin comparare cu limitele de control prevăzute în normativele de calitate.

Efectuînd astfel controlul calității loturilor de rețele cord, s-au ivit situații cînd loturi încadrate de furnizor la calitatea I să fie încadrate de beneficiar la calitatea II.

Această situație s-a produs la caracteristica de calitate „sarcina de rupere" (rezistența la rupere), caracteristică calitativă principală a rețelelor cord.

La primele cazuri semnalate, această neconcordanță a valorilor medii obținute la furnizor și beneficiar a fost pusă pe seama neuniformității firelor.

Scanned with OKEN Scanner

Din punct de vedere statistic, aceasta s-ar întîmpla atunci cînd calitatea medie are o valoare foarte apropiată de limita de control şi deci probabilitatea ca un lot bun să fie respins ar fi mai mare decît riscul ales α.

La un moment dat, fiind înregistrate mai multe asemenea situaţii, s-a pus problema determinării cauzelor reale care duc la astfel de situaţii, mai ales că rezultatele obţinute în laboratoarele furnizorului nu sugerau totuşi o apropiere evidentă a calităţii medii faţă de limita de control.

Pentru început, s-au extras unele loturi, determinîndu-se valorile sarcinii la rupere pentru două sortimente cord, atît de la furnizor, cît şi de la beneficiar.

Aceste date arătau o neconcordanţă de valori, în general, la toate loturile.

Din analiza graficelor, nu s-a putut trage concluzia dacă aceste diferenţe sînt întîmplătoare sau dacă sînt semnificative.

Ne propunem mai întîi să eliminăm din datele obţinute acele valori care diferă semnificativ de tendinţa generală a datelor.

Trebuie să specificăm faptul că vom studia modul în care evoluează diferenţa dintre valorile rezistenţei la rupere obţinute la furnizor şi beneficiar.

Utilizînd testul lui Dixon, s-au alcătuit tabele din care pe baza calculelor să poată fi eliminate valorile ce se abat semnificativ.

În aceste tabele x_i reprezintă valorile rezistenţei la rupere obţinute la furnizor, y_i valorile corespunzătoare obţinute la beneficiar, iar d_i reprezintă diferenţa dintre perechile $(x_i,\ y_i) : d_i = x_i - y_i$.

Calculele arată că:

$$\Sigma R = 1{,}14 + 1{,}02 = 2{,}16, \quad Q_1 = \frac{1{,}14}{2{,}16} = 0{,}53, \quad Q_2 = \frac{1{,}02}{2{,}16} = 0{,}47,$$

iar din tabelul testului Dixon pentru $k = 2$ grupe şi $n = 9$ date se obţine: $Q_t = 0{,}37$. Deoarece $Q_1 > Q_t$ şi $Q_2 > Q_t$ trebuie eliminate valorile — 0,52, 0,62, 1,07 şi 0,05 (ele au fost încadrate în tabelul 4.4).

Etapa următoare eliminării valorilor care se abat semnificativ este cea de a studia dacă rezultatele obţinute la controlul caracteristicii „rezistenţă la rupere“ pe acelaşi lot la furnizor şi respectiv la beneficiar diferă sau nu semnificativ.

Scanned with OKEN Scanner

Tabelul 4.4

Eliminarea valorilor extreme pentru sortimentul I cu testul lui Dixon

Nr.	1			2		
	x_i	y_i	d_i	x_i	y_i	d_i
1	12,54	13,06	[—0,52]	12,82	12,50	0,32
2	12,54	12,50	0,04	12,73	12,60	0,13
3	13,12	12,50	[0,62]	13,13	12,06	[1,07]
4	12,96	12,89	0,07	12,73	12,50	0,23
5	12,89	12,70	0,19	12,88	12,50	0,38
6	12,80	12,70	0,10	12,73	12,60	0,13
7	12,76	12,60	0,16	12,85	12,40	0,45
8	13,00	12,50	0,50	12,84	12,50	0,34
9	12,87	12,50	0,37	12,65	12,60	[0,05]
R			1,14			1,02

Tabelul 4.5

Eliminarea valorilor extreme pentru sortimentul II cu testul lui Dixon

Nr.	1			2		
	x_i	y_i	d_i	x_i	y_i	d_i
1	21,50	21,00	0,50	21,50	21,20	0,30
2	21,50	21,30	0,20	21,80	21,10	0,70
3	21,50	21,00	0,50	21,60	21,40	0,20
4	21,60	21,40	0,20	21,30	21,10	0,20
5	21,60	21,10	0,50	21,20	20,10	1,10
6	21,80	21,20	[0,60]	21,50	21,20	0,30
7	21,30	21,60	—0,30	21,50	21,20	0,30
8	21,50	21,30	0,20	21,30	21,00	0,30
9	21,50	21,10	0,40	21,30	21,00	0,30
R			0,90			0,90

Scanned with OKEN Scanner

Din punct de vedere statistic aceasta revine la verificarea ipotezei statistice:

$$H_0: \bar{d} = 0$$

față de alternativa

$$H_1: \bar{d} \neq 0$$

unde $\bar{d}$ reprezintă media de selecție a diferențelor d_i.

Pentru verificarea acestei ipoteze statistice se folosește statistica (4.8) (testul t).

În cazul nostru $\bar{X} = \bar{d}$, iar $\mu_0 = 0$. Pe baza datelor rămase în urma eliminării s-a calculat pentru fiecare sortiment statistica t.

Avem deci

$$n = 14, \quad \nu = n - 1 = 13, \quad \Sigma d_i = 3{,}41, \quad \bar{d} = 0{,}24;$$

$$S_d = 0{,}15, \quad S_{\bar{d}} = \frac{S_d}{\sqrt{n}} = \frac{0{,}15}{3{,}74} = 0{,}04.$$

Se obține astfel

$$t = \frac{\bar{d}}{S_{\bar{d}}} = \frac{0{,}24}{0{,}04} = 6.$$

Din tabelul 6, găsim pentru $\nu = 13$ și un risc $\alpha = 0{,}05$, valoarea $t_{13,0,05} = 2{,}16$.

Tabelul 4.6

Date folosite pentru calculul statisticii „t" pentru sortimentul I

Nr.	Diferențele între valorile obținute		
1	0,04	8	0,32
2	0,07	9	0,13
3	0,19	10	0,23
4	0,10	11	0,38
5	0,16	12	0,13
6	0,50	13	0,45
7	0,37	14	0,34

Deoarece $t = 6 > t_{13,0,05} = 2{,}16$, ipoteza se respinge.

Scanned with OKEN Scanner

Tabelul 4.7

Date folosite pentru calculul statisticii „t" pentru sortimentul II

Nr.	Diferențele între valorile obținute		
1	0,50	8	0,30
2	0,20	9	0,70
3	0,50	10	0,20
4	0,20	11	0,30
5	0,50	12	0,30
6	0,20	13	0,40
7	0,40	14	0,40

Avem

$$n = 14, \quad \nu = 13, \quad \Sigma d_i = 5{,}10, \quad \bar{d} = 0{,}36;$$

$$S_d = 0{,}15, \quad S_{\bar{d}} = \frac{S_d}{\sqrt{n}} = 0{,}04,$$

de unde

$$t = \frac{\bar{d}}{S_{\bar{d}}} = \frac{0{,}36}{0{,}04} = 9.$$

Deoarece

$t = 9 > t_{13,0,05} = 2{,}16$ și în acest caz ipoteza se respinge.

Așadar, pînă acum, prin respingerea ipotezei $H_0 : \bar{d} = 0$ putem trage concluzia că, rezultatele obținute pe același lot la furnizor și beneficiar diferă semnificativ.

După cum a reieșit pînă acum, între măsurătorile efectuate asupra rețelelor cord din fire relon privitor la caracteristica „rezistență la rupere", furnizor și beneficiar există o diferență semnificativă pusă în evidență tocmai cu ajutorul testului t.

O problemă importantă deschisă de această analiză este aceea a stabilirii cauzelor care conduc la această situație.

În cele ce urmează se vor verifica unele ipoteze emise în acest sens, cu ajutorul aparatului statistico-matematic descris în paragrafele anterioare.

Scanned with OKEN Scanner

Datorită faptului că între metoda de determinare a rezistenței practicată la furnizor (care constă în determinarea acesteia pe epruvetă condiționată) și cea de la beneficiar (unde determinările se execută pe fir uscat) a apărut în mod natural ca o primă ipoteză neconcordanța datorată acestor două metode diferite.

Pentru verificarea acestei ipoteze s-a procedat după cum urmează: s-a luat jumătatea de metru de probă care inițial a fost trimisă beneficiarului însoțind lotul livrat și din care s-au extras probe ce au fost analizate în laboratoarele acestuia, obținîndu-se valorile de selecție (x_{1i}), $i = 1,...,20$ și s-au extras probe ce au fost analizate în laboratoarele furnizorului, obținîndu-se valorile de selecție (x_{2j}) $j = 1,...,20$.

S-a procedat și invers, adică din jumătatea de metru de probă analizată întîi la furnizor, obținîndu-se valorile de selecție (y_{1k}) $k = 1,...,20$ s-au extras probe care s-au analizat la beneficiar, obținîndu-se valorile de selecție (y_{2l}) $l = 1,...,20$.

Aceste operații au fost făcute cu scopul de a se urmări dacă în această situație există sau nu diferențe semnificative.

Este important de accentuat că probele asupra cărora s-au făcut analizele au fost supuse acelorași condiții după cum urmează:

a) proba din care au rezultat (y_{1k}) și (y_{2l}) a fost detașată de lot înaintea livrării acestuia;

b) proba din care au rezultat (x_{1i}) și (x_{2j}) a fost extrasă din lot (respectiv dintr-o rolă a acestuia) după ce a trecut împreună cu acesta prin întreaga filieră a livrării, transportului și recepționării.

Din punct de vedere statistic, ipoteza enunțată mai sus se formalizează în felul următor:

Trebuie să se verifice ipoteza statistică

$H_0 : \mu_1 = \mu_2$ față de alternativa $H_0' : \mu_1 \neq \mu_2$ precum și ipoteza statistică

$H_1 : \mu_1' = \mu_2'$ față de alternativa $H_1' : \mu_1' \neq \mu_2'$ unde μ_1, μ_2, μ_1', μ_2' sînt mediile populațiilor normale din care provin selecțiile x_{1i}, x_{2j}, y_{1k} respectiv y_{2l}.

Pentru verificarea acestor ipoteze s-a folosit testul t după ce în prealabil s-a verificat dacă dispersiile respective diferă sau nu semnificativ, pentru cele două sortimente de rețele cord analizate.

Se va exemplifica modul în care s-au verificat ipotezele respective pe baza rezultatelor măsurătorilor efectuate asupra me-

Scanned with OKEN Scanner

trului de probă extras înainte de livrare pentru sortimentul II de rețele cord din fire relon.

Calculele parametrilor statistici necesari verificării ipotezei sînt sintetizate în tabelele 4.8 și 4.9. În aceste tabele y_{1k} reprezintă valorile de selecție obținute la furnizor iar y_{2l} valorile obținute la beneficiar.

Tabelul 4.8

nr.	y_{1k}	$y_{1k} - \bar{y}_1$	$(y_{1k} - \bar{y}_1)^2$
1	21,6	0,38	0,1444
2	21,2	—0,02	0,0004
3	21,8	0,58	0,3364
4	21,0	—0,22	0,0484
5	21,5	+0,28	0,0784
6	20,9	—0,32	0,1024
7	21,1	—0,12	0,0144
8	21,8	0,58	0,3364
9	21,1	—0,12	0,0144
10	21,1	—0,12	0,0144
11	20,1	—1,12	1,2544
12	21,5	0,28	0,0784
13	21,2	—0,02	0,0004
14	21,0	—0,22	0,0484
15	21,1	—0,12	0,0144
16	21,2	—0,02	0,0004
17	21,3	—0,08	0,0064
18	21,6	0,38	0,1444
19	21,5	0,28	0,0784
20	20,8	—0,42	0,1764
	424,4		2,8920

Folosind datele din tabelul 4.8, rezultă:

$$\bar{y}_1 = \frac{1}{n}\sum_1^n y_{1k} = \frac{1}{20} \cdot 424{,}4 = 21{,}22,$$

$$S_1^2 = \frac{1}{n-1}\sum_1^n (y_{1k} - \bar{y}_1)^2 = \frac{1}{19} \cdot 2{,}892 = 0{,}1522,$$

Scanned with OKEN Scanner

Tabelul 4.9

nr.	y_{2i}	$(y_{2i} - \bar{y}_2)^2$	$(y_{2i} - \bar{y}_2)^2$
1	21,5	0,24	0,0576
2	21,6	0,34	0,1156
3	21,9	0,64	0,4096
4	20,7	—0,56	0,3136
5	21,1	—0,16	0,0256
6	21,2	—0,06	0,0036
7	21,1	—0,16	0,0256
8	20,8	—0,46	0,2116
9	21,0	—0,26	0,0676
10	21,4	0,14	0,0196
11	21,0	—0,26	0,0676
12	21,5	0,24	0,0576
13	21,3	0,04	0,0015
14	21,2	—0,06	0,0036
15	21,1	—0,16	0,0256
16	21,1	—0,16	0,0256
17	21,4	0,14	0,0196
18	21,5	0,24	0,0576
19	21,5	0,24	0,0576
20	21,4	0,14	0,0196
	425,3		1,5860

Folosind datele din tabelul 4.9 rezultă:

$$\bar{y}_2 = \frac{1}{n}\sum_1^n y_{2i} = \frac{1}{20} \cdot 425{,}3 = 21{,}26,$$

$$S_2^2 = \frac{1}{n-1}\sum_1^n (y_{2i} - \bar{y}_2)^2 = \frac{1}{19} \cdot 1{,}5860 = 0{,}0834.$$

Pentru aplicarea testului t s-a verificat mai întîi dacă are loc egalitatea dispersiilor.

Pentru aceasta, se aplică testul F și anume

$$F = \frac{S_1^2}{S_2^2} = \frac{0{,}1522}{0{,}0834} = 1{,}8249.$$

Scanned with OKEN Scanner

Din tabelul 5a găsim $F^{*}_{0,05;\,19,\,19} \simeq 2{,}15$.

Cum $F < F_{0,05;\,19,\,19}$ tragem concluzia că cele două dispersii sînt practic egale.

În această situaţie, statistica (4.22) este

$$t = \frac{\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(|\bar{y}_1 - \bar{y}_2|)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}} \tag{4.22'}$$

cu $n_1 = n_2 = 20$ şi ţinînd seama de situaţia concretă

$$t = \frac{20 \times 0{,}04 \times 0{,}581}{\sqrt{\frac{5}{19} \times 0{,}2356}} = \frac{0{,}1264}{0{,}2489} = 0{,}5078.$$

Din tabelul 6 rezultă $t_{0,05;\,38} = 2{,}02$.

Deoarece $t < t_{0,05;\,38}$ se acceptă ipoteza egalităţii mediilor.

În mod analog s-a procedat şi pentru celelalte situaţii atît în cazul sortimentului II cît şi pentru I.

În toate cazurile concluziile analizei statistice au fost aceleaşi cu cea prezentată în exemplul de mai sus.

Acest fapt permite afirmaţia că nu metodele diferite de măsurătoare sînt cauze care duc la neconcordanţe, deoarece asupra eşantioanelor care au beneficiat de aceleaşi condiţii, aplicarea respectivelor metode de măsurătoare au condus la rezultate care nu diferă semnificativ din punct de vedere statistic.

Eliminînd deci ipoteza emisă anterior se pune în mod natural problema găsirii adevăratelor cauze ce conduc la neconcordanţe.

În această ordine de idei, s-a făcut ipoteza că solicitările suportate de părţile exterioare ale rolelor de reţele cord — părţi în care se află şi metrul de probă — în timpul tuturor manevrelor efectuate cu acestea pînă cînd ele ajung la beneficiar, ar fi o cauză posibilă.

Pentru verificarea acestei ipoteze s-au iniţiat trei tipuri de experimente şi anume:

Experimentul 1: S-au recoltat probe din reţele cord existente la beneficiar din exteriorul rolei şi la diferite nivele către interiorul rolei.

---

* vezi [18] tabela VIII, pag. 407.

Scanned with OKEN Scanner

Pentru fiecare din aceste eşantioane s-au calculat estimaţiile mediei şi dispersiei, urmărind verificarea ipotezei statistice dacă valorile medii la diferite nivele diferă sau nu semnificativ între ele. Calculele sînt sintetizate în tabelul 4.10 pentru sortimentul I de reţele cord.

Tabelul 4.10

nr.	x_i	y_i	z_i
1	12,250	11,750	13,050
2	12,650	12,450	13,150
3	11,950	13,500	12,850
4	11,900	12,500	12,500
5	11,750	12,300	12,750
6	12,350	12,800	13,200
7	12,600	12,550	12,850
8	12,300	12,950	12,800
9	12,500	11,950	13,350
10	12,100	12,100	12,250
11	12,150	12,400	13,100
12	11,850	13,500	12,650
13	12,800	12,600	12,950
14	11,800	12,650	13,350
15	11,600	11,300	12,700

$$\bar{x} = 12{,}170, \quad \bar{y} = 12{,}486, \quad \bar{z} = 12{,}900$$

$$S_x^2 = 0{,}1316, \quad S_y^2 = 0{,}3456, \quad S_z^2 = 0{,}096$$

unde x_i reprezintă valorile rezistenţei măsurată pe eşantionul extras la capătul rolei,

y_i reprezintă valorile rezistenţei măsurată la 10 m de capătul rolei,

z_i reprezintă valorile rezistenţei măsurată pe eşantionul extras la 50 m de capătul rolei.

a) Se verifică mai întîi ipoteza statistică:

$$H_0: \mu_1 = \mu_2 \text{ faţă de alternativa } H_1: \mu_1 \neq \mu_2$$

unde μ_1, μ_2 au semnificaţia de medii ale populaţiilor respective.

Pentru aceasta se foloseşte mai întîi testul F pentru a vedea dacă între dispersii există sau nu diferenţe semnificative.

Scanned with OKEN Scanner

Avem

$$F = \frac{S_y^2}{S_x^2} = \frac{0,3456}{0,1316} = 2,6261.$$

Ţinînd cont că $n_1 = n_2 = 15$ din tabele se găseşte, pentru riscul α ales 0,05,

$$F_{0,05;14,14} = 2,48.$$

Cum $F > F_{0,05,14,14}$ tragem concluzia că dispersiile sînt diferite.

În această situaţie verificarea ipotezei

$$H_0: \mu_1 = \mu_2 \text{ în condiţia } \sigma_1^2 \neq \sigma_2^2$$

se face cu ajutorul statisticii (4.24), care are o repartiţie Student cu numărul de grade de libertate date de (4.33).

Pentru cazul de faţă obţinem

$$t = \frac{12,310 - 12,170}{\sqrt{\frac{0,1316}{15} + \frac{0,3456}{15}}} = \frac{0,1400}{0,1707} = 0,8206$$

iar din (4.32) obţinem

$$c = \frac{\frac{0,1316}{15}}{\frac{0,1316}{15} + \frac{0,3456}{15}} = \frac{0,1316}{0,4772} = 0,2757$$

de unde $1-c = 0,7243$, $c^2 = 0,0760$ şi $(1-c)^2 = 0,5246$. Numărul de grade de libertate este, conform cu (4.33),

$$\frac{1}{\frac{0,0760}{14} + \frac{0,5246}{14}} = \frac{14}{0,6006} = 23,3 \simeq 23.$$

Din tabelul 6 pentru riscul de 0,05 şi 23 grade de libertate găsim:

$$t_{0,05;23} = 2,069.$$

Cum $t < t_{0,05;23}$ rezultă că se acceptă ipoteza egalităţii mediilor.

Scanned with OKEN Scanner

b) Se va verifica acum ipoteza statistică:

$$H_1: \bar{Y} = \bar{Z} \text{ față de alternativa } H_1': \bar{Y} \neq \bar{Z}$$

obținem

$$F = \frac{S_y^2}{S_z^2} = \frac{0,3456}{0,1031} = 3,3520$$

iar din tabelul 5a, $F_{0,05;\,14,\,14} = 2,48$.
Deoarece $F > F_{0,05;\,14,\,14}$ și în acest caz ipoteza egalității dispersiilor se respinge.

Statistica t, calculată ca în cazul precedent, are valoarea

$$t = \frac{12,900 - 12,486}{\sqrt{\frac{0,3456}{15} + \frac{0,1031}{15}}} = \frac{0,4140}{0,1726} = 2,3985;$$

celelalte elemente sînt calculate cu relațiile (4.33) și (4.32): $c = 0,7702$, $c^2 = 0,5932$, $1 - c = 0,2298$, $(1 - c)^2 = 0,0528$ și $n^* \simeq 20$. Din tabelul 6

$$t_{0,05;\,20} = 2,086.$$

Se vede că $t > t_{0,05;\,20}$ deci ipoteza egalității mediilor se respinge în acest caz.

c) Se va verifica ipoteza statistică

$$H_2: \bar{X} = \bar{Z} \text{ față de alternativa } H_2': \bar{X} \neq \bar{Z}.$$

Testul F furnizează rezultatul

$$F = \frac{S_x^2}{S_z^2} = \frac{0,1316}{0,1031} = 1,2764.$$

Cum $F_{0,05;\,14,\,14} = 2,48$ și $F < F_{0,05;\,14,\,14}$ dispersiile se consideră practic egale.

Conform relației (4.22′) statistica t are valoarea

$$t = \frac{\sqrt{\frac{15.15}{15+15}}\,(12,900 - 12,470)}{\sqrt{\frac{15.0,1316 + 15.0,1031}{15+15-2}}} = \frac{1,9885}{0,3545} = 5,6093.$$

Scanned with OKEN Scanner

Din tabelul 6

$$t_{0,05;\,28} = 2{,}048.$$

Cum $t > t_{0,05;\,28}$ ipoteza egalității mediilor se respinge.

Aceeași procedură a fost aplicată și pentru sortimentul II de rețele cord.

S-a extras și analizat eșantionul (u_i) cu $i = 1, \ldots, 18$ de la distanța de 10 m de la exterior către interiorul rolei și eșantionul (v_i), $i = 1, \ldots, 18$, distanța fiind de data aceasta de 30 m, măsurată în același mod.

Se vor da aici doar rezultatele obținute, întrucît calculul este analog.

Trebuie verificată ipoteza statistică

$$H_3: \bar{U} = \bar{V} \text{ față de alternativa } H_3': \bar{U} \neq \bar{V}.$$

Estimațiile parametrilor statistici sînt:
$\bar{u} = 21{,}313$ cu dispersia $S_u^2 = 0{,}1257$ și $\bar{v} = 21{,}745$ cu dispersia $S_v^2 = 0{,}4056$ unde $n_1 = n_2 = 18$.
S-a obținut

$$F = \frac{S_v^2}{S_u^2} = \frac{0{,}4056}{0{,}1257} = 3{,}2267, \text{ iar din tabelul 5a}$$

$$F_{0,05;\,17,17} = 2{,}29.$$

Cum $F > F_{0,05,\,17,17}$ ipoteza egalității dispersiilor s-a respins; folosind relațiile (4.24), (4.33), (4.32) avem:

$$t = \frac{21{,}745 - 21{,}313}{\sqrt{\frac{0{,}1257}{18} + \frac{0{,}4056}{18}}} = \frac{0{,}4320}{0{,}1717} = 2{,}5160.$$

Celelalte elemente sînt: $c = 0{,}365$, $c^2 = 0{,}0559$, $1 - c = 0{,}7635$, $(1 - c)^2 = 0{,}5829$ și $n^* = 26$.

Din tabele, $t_{0,05;\,26} = 2{,}056$.

Deoarece $t > t_{0,05,\,26}$ ipoteza egalității mediilor se respinge.

Drept concluzie a rezultatelor experimentului, interpretate cu ajutorul metodelor statistico-matematice, se poate spune că solicitările mecanice din timpul transportului au ca efect scăderea valorii rezistenței la exteriorul rolelor, scădere care se diminuează ca valoare odată cu pătrunderea în adîncimea rolei.

Scanned with OKEN Scanner

În acest sens valoarea medie obținută pe eșantioanele extrase de la 50 m spre interior la cordul I respectiv la 3 m pe cordul II sînt practic egale cu valorile medii obținute în urma analizei jumătății de metru de probă efectuată la furnizor înainte de livrare.

Să remarcăm faptul că deciziile luate cu ajutorul metodelor statistico-matematice, metode bine fundamentate din punct de vedere științific cu un înalt grad de justețe, aceste metode venind în ajutorul experimentatorilor ferindu-i de a lua decizii în mod empiric care de foarte multe ori se dovedesc a fi false.

Este semnificativ chiar în cazul de față faptul că la o examinare superficială a mediilor $\bar{x}$ și $\bar{y}$ s-ar putea considera că acestea diferă semnificativ ori aplicarea testului t conduce la respingerea acestei ipoteze.

Experimentul 2. Acest experiment a fost organizat în laboratorul furnizorului luîndu-se jumătatea de metru de probă existentă din care s-a extras un eșantion măsurîndu-se rezistența și determinîndu-se valorile de selecție (ξ_{1i}) $i = 1, \ldots, 20$.

După aceasta, aceeași jumătate de metru de probă s-a supus în laborator unor solicitări asemănătoare cu acelea la care sînt supuse rolele de rețele cord în timpul manipulării și transportului. După aceasta, s-au făcut din nou măsurători asupra rezistenței, obținîndu-se valorile de selecție: (ξ_{2j}), $j = 1, \ldots, 20$ pentru cordul $840/140 \times 2$. Cu ajutorul acestor valori s-au calculat estimațiile.

$$\bar{\xi}_1 = 13{,}105 \text{ cu } S_1^2 = 0{,}2583, \quad \bar{\xi}_2 = 12{,}285 \text{ cu } S_2^2 = 0{,}3145.$$

S-a propus verificarea ipotezei statistice

$$H_4: M_1 = M_2 \text{ față de alternativa } H_4': M_1 \neq M_2.$$

unde M_1 și M_2 au semnificația mediilor populațiilor normale din care s-au extras selecțiile (ξ_{1i}) și (ξ_{2j}).

Urmînd tehnica de calcul aplicată pînă acum, găsim

$$F = \frac{S_2^2}{S_1^2} = \frac{0{,}3145}{0{,}2585} = 1{,}2175.$$

Scanned with OKEN Scanner

Din tabele $F_{0,05;\,19,19} = 2,12$. Cum $F < F_{0,05;\,19,19}$, ipoteza egalității dispersiilor se acceptă. În acest caz statistica t se calculează cu relația (4.22′).

$$t = \frac{\sqrt{\frac{20 \cdot 20}{20 + 20}}\,(13,105 - 12,285)}{\sqrt{\frac{20 \cdot 0,2583 + 20 \cdot 0,3145}{20 + 20 - 2}}} = \frac{2,5928}{0,5489} = 4,7236.$$

Din tabele se găsește $t_{0,05;\,38} = 2,02$.

Cum $t > t_{0,05;\,38}$ ipoteza egalității mediilor se respinge. Pentru sortimentul II în cadrul aceluiași experiment s-au obținut înainte de solicitare valorile (η_{1i}) $i = 1, 2, \ldots, 20$ iar după solicitare valorile (η_{2j}) $j = 1, \ldots, 20$ cu ajutorul cărora s-au găsit următoarele estimații ale parametrilor:

$$\bar{\eta}_1 = 21,350 \quad \text{cu} \quad S_1^2 = 0,2422 \quad \text{și} \quad \bar{\eta}_2 = 20,040 \quad \text{cu} \quad S_2^2 = 0,3792.$$

S-a propus verificarea ipotezei statistice:

$$H_5: M_1' = M_2' \text{ față de alternativa } H_5': M_1' \neq M_2'$$

unde M_1' și M_2' au aceeași semnificație ca și M_1, M_2.

Calculele au furnizat rezultatele:

$$F = 1,5656 < F_{0,05;\,19,19} = 2,12$$

deci se acceptă ipoteza egalității dispersiilor.

Din relația (4.22′) s-a găsit $t = 7,2390$. Deoarece

$$t = 7,2390 > t_{0,05;\,38} = 2,02,$$

și pentru acest sortiment ipoteza egalității mediilor s-a respins.

Drept concluzie la acest experiment se poate spune că rezistența a scăzut în mod semnificativ după solicitările la care au fost supuse firele de rețele cord în laborator.

Experimentul 3. Acest experiment constă în faptul că la o rolă de rețea cord s-au țesut 2 m de probă: un metru de probă s-a detașat imediat după țesere și din el s-a extras un eșantion la care s-a măsurat rezistența obținîndu-se valorile (ω_{1i}) $i = 1, \ldots, 20$. Celălalt metru de probă a rămas pe rolă. După aceasta s-a

Scanned with OKEN Scanner

manipulat rola respectivă simulîndu-se drumul parcurs pînă la beneficiar și anume:

a) în secție: ambalare, cîntărire;

b) în magazia de produse finite: stocare și încărcare în vagon.

c) la beneficiar: descărcare, transport și depozitare.

După această manipulare s-a detașat și cel de-al doilea metru de probă care s-a analizat în laborator obținîndu-se pentru rezistență valorile (ω_{2j}), $j = 1, ..., 20$.

S-a propus verificarea ipotezei statistice:

$$H_6: \Omega_1 = \Omega_2 \text{ față de alternativa } H_6': \Omega_1 \neq \Omega_2$$

unde Ω_1 și Ω_2 au semnificația de medii de populații.

Estimațiile parametrilor sînt următoarele:

$$\overline{\omega}_1 = 21{,}596 \text{ cu } S_1^2 = 0{,}3241 \text{ și } \overline{\omega}_2 = 20{,}766 \text{ cu } S_2^2 = 0{,}4722.$$

Deoarece

$$F = 1{,}4569 < F_{0{,}05;\,19,19} = 2{,}12$$

se acceptă ipoteza egalității dispersiilor.

Din relația (4.22′) și din tabelul 6 rezultă

$$t = 4{,}0543 > t_{0{,}05;\,38} = 2{,}02.$$

Așadar și în acest caz ipoteza egalității mediilor se respinge.

Concluzia care se trage este că și în cazul acestui experiment solicitările rețelelor cord conduc la scăderea rezistenței.

§ 6. Forme analoage ale testului t-necentrat

Fiind dată o populație normală $N(\mu,\sigma^2)$ o constantă reală T și un număr p, $(0 < p < 1)$ trebuie să se verifice ipoteza statistică:

$$\int_T^\infty \varphi(x;\, \mu,\, \sigma^2)\, dx \leqslant p$$

unde

$$\varphi(x;\, \mu,\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Fie ξ_a o variabilă aleatoare normală cu densitatea $\varphi(x, a, 1)$ și η o variabilă aleatoare nenegativă, stochastic independentă de ξ_a cu funcția de repartiție continuă $F(x)$.

Scanned with OKEN Scanner

Se notează prin $K(x; a; \eta)$ funcția de repartiție a raportului ξ_a/η. Se demonstrează că:

$$K(x, a; \eta) = \begin{cases} 1 - x\int_0^\infty F(y)\,\varphi(xy - a;\ 0{,}1)\,dy & \text{pentru } x > 0, \\ -x\int_0^\infty F(y)\,\varphi(xy - a;\ 0;\ 1)\,dy & \text{pentru } x < 0. \end{cases}$$

Repartițiile cu funcția de repartiție de forma de mai sus sînt numite, de către Fr. Zítek, repartiții de tip U.

Fie $x_1, \ldots, x_n$ o selecție de volum n dintr-o populație normală $N(\mu, \sigma^2)$. Se notează:

$\bar{x} = \frac{1}{n}\sum_1^n x_i$ — media de selecție,

$S_d = \sqrt{\frac{1}{n}\sum_1^n (x_i - \bar{x})^2}$ — estimatorul deplasat al abaterii standard (σ),

$\Delta = \frac{1}{n}\sum_1^n |x_i - \bar{x}|$ — abaterea medie absolută,

$R = x_{(n)} - x_{(1)}$ — amplitudinea de selecție

unde $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ reprezintă șirul ordonat al valorilor de selecție.

Fie $S = S(x_1, \ldots, x_n)$ o funcție de selecție cu următoarele proprietăți:

a) este nenegativă pentru orice variabilă de selecție x_i;

b) admite o funcție de repartiție continuă;

c) pentru orice „a" real este invariantă la translații de tipul $(x_i + a)$, adică:

$$S(x_1 + a, \ldots, x_n + a) = S(x_1, \ldots, x_n).$$

În aceste condiții se definește statistica:

$$U_p = \sqrt{n}\,\frac{\mu + l_p\sigma - \bar{x}}{S}$$

unde $0 < p < 1$, iar

$$p = \frac{1}{\sqrt{2\pi}}\int_{l_p}^\infty \exp\left(-\frac{x^2}{2}\right)dx.$$

Scanned with OKEN Scanner

Punînd U_p sub forma:

$$U_p = \frac{\sqrt{n} \cdot \frac{\mu - \bar{x}}{\sigma} + l_p \sqrt{n}}{\frac{S}{\sigma}}$$

atunci se observă că U_p are funcţia de repartiţie:

$$K\left(x;\ l_p \sqrt{n};\ \frac{S}{\sigma}\right).$$

Cazuri particulare importante:

1) $S = S_d$ — se obţine testul t — necentrat (Johnson-Welch);

2) $S = R$ — se obţine testul Lord (E. Lord, în 1947, a dat o formă modificată a testului t folosind amplitudinea de selecţie în locul abaterii standard de selecţie);

3) $p = 0{,}5$ $(l_p = 0)$ şi $S = S_d$ — se obţine

$$U_{0,5} = t = \sqrt{n} \cdot \frac{\mu - \bar{x}}{S_d}$$

care reprezintă testul clasic al lui Student.

Fr. Zítek propune statistica

$$Z = \frac{T - \bar{x}}{\Delta}$$

pentru verificarea ipotezei propuse. Zítek a tabelat statistica Z pentru p $= 0{,}05$ $(0{,}025)$ şi $\alpha = 0{,}05;\ 0{,}025;\ 0{,}01$ $(n = 5,...,10)$.

Valorile critice pentru statistica Z sînt date în tabelul 4.11.

Tabelul 4.11

l_p	1,645			1,960		
n \ α	0,01	0,025	0,05	0,01	0,025	0,05
5	0,709	0,895	1,069	1,005	1,200	1,385
6	0,795	0,971	1,130	1,097	1,277	1,448
7	0,870	1,033	1,181	1,167	1,339	1,500
8	0,928	1,084	1,225	1,225	1,390	1,544
9	0,977	1,126	1,262	1,276	1,434	1,582
10	1,020	1,161	1,293	1,320	1,473	1,616

Scanned with OKEN Scanner

Exemplu. (V. Gh. Vodă; *Aplicații ale unui criteriu statistic în chimie*, „Revista de chimie" Nr. 1/1972). La determinarea cuprului prin electroliză internă s-au obținut următoarele date (în miligrame)

i	1	2	3	4	5	6	7	8
$x_{(i)}$	49,53	49,80	49,81	49,81	49,82	49,92	49,98	50,01

Obținem $\bar{x} = 49{,}84$. Alegem $T = 50{,}00$ și ne propunem verificarea ipotezei statistice:

$$H: \frac{1}{\sqrt{2\pi}} \int_{50{,}00}^{+\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \leqslant 0{,}05.$$

Valoarea statisticii lui Zítek, ținînd seama că $\Delta = 0{,}11375$, este

$$Z = \frac{50{,}00 - 49{,}84}{0{,}11375} = \frac{0{,}16000}{0{,}11375} \simeq 1{,}4.$$

Din tabelul 4.11 găsim $Z_{l_p;\alpha;n} = Z_{1{,}645;\,0{,}05;\,8} = 1{,}225$. Cum $Z_{calculat} > Z_{tabelat}$ ipoteza se acceptă. Valoarea T are aici semnificația unei toleranțe superioare.

Scanned with OKEN Scanner

CAPITOLUL
5

TESTE NEPARAMETRICE

§ 1. Introducere

După cum am văzut o mare parte a tehnicilor statistice se referă la variabile continui repartizate normal. Această restricție nu era esențială deoarece:

a) adeseori se poate găsi o transformare a variabilei, în presupunerea că i se cunoaște repartiția, care să conducă la o variabilă repartizată aproximativ normal, și

b) teorema limită centrală ne spune că dacă variabila care caracterizează populația are dispersia finită, atunci repartiția mediei de selecție, pentru $n \to \infty$, tinde către repartiția normală.

Importanța presupunerii de normalitate depinde de tehnicile statistice ce trebuie folosite, de natura neconcordanței cu normalitatea și de scopul final al analizei.

În ultimul timp s-au dezvoltat tehnici care nu fac presupuneri asupra formei repartiției însă se presupune că repartiția este continuă. Aceste metode sînt cunoscute sub numele de *metode neparametrice* sau *metoda repartițiilor libere.*

§ 2. Testul semnelor

Să presupunem că vectorul aleator (x, y) are densitatea de repartiție $f(x, y)$ și că $f(x, y) = f(y, x)$. Presupunerea făcută atrage după sine simetria funcției $f(x' y)$ față de dreapta $y - x = 0$.

Avem

$$P(y < x) = P(x < y) = \frac{1}{2}$$

Scanned with OKEN Scanner

de unde

$$P(y - x < 0) = P(y - x > 0) = \frac{1}{2}. \tag{2.1}$$

Dacă punem $z = y - x$, ținînd seama de (2.1), obținem

$$P(z < 0) = P(z > 0) = \frac{1}{2}$$

ceea ce arată că variabila z are drept mediană pe zero și o repartiție simetrică.

Pentru testul semnelor se cere numai ca mediana repartiției să fie zero.

Fie

$$Z_i = \begin{cases} 1 \text{ dacă } z_i > 0, \\ 0 \text{ dacă } z_i < 0. \end{cases}$$

Presupunem că densitatea de repatiție $f(x, y)$ este continuă în origine, ceea ce atrage după sine continuitatea repartiției lui z. Urmează că

$$P(z_i = 0) = 0.$$

Variabilele z_i sînt independente astfel încît sîntem în cazul unei experiențe binomiale în care în fiecare experiență probabilitatea succesului este $\frac{1}{2}\left[P(Z_i = 1) = \frac{1}{2}\right]$. Urmează că $\sum_{i=1}^{n} Z_i$ este repartizată binomial $b\left(n_{i,,}, \frac{1}{2}\right)$.

O ipoteză alternativă este

$$x_i' = x_i - d_i, \qquad d_i > 0$$

și în acest caz graficul densității de repartiție corespunzător variabilelor x' și y se deplasează spre stînga și

$$P(z_i > 0) = P(x_i - x_i' > 0) = P(y_i > x_i') > \frac{1}{2}.$$

Scanned with OKEN Scanner

$P(z_i > 0) > \frac{1}{2}$ nu este constantă, în mod necesar, pentru diferitele valori ale lui i, și astfel $\sum_{i=1}^{n} z_i$ nu are, în mod necesar, o repartiție binomială.

Dacă $\sum_{i=1}^{n} z_i$ ia o valoare mai mare decît $\frac{n}{2}$, atunci o regiune critică acceptabilă, va fi dată de coada superioară.

Astfel testul semnelor, cînd avem m diferențe pozitive într-o selecție de volum n, dă P-cuantila

$$P\left(\sum_{1}^{n} z_i \geqslant m \mid n, \theta = \frac{1}{2}\right) = \sum_{k=m}^{n} C_n^k \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k} = \frac{1}{2^n} \sum_{k=m}^{n} C_n^k.$$

Pentru valori mari ale lui n și m se folosește aproximația normală dată de (2.3.3).

Pînă acum am presupus că în ipoteza nulă mediana diferenței $x_i - y_i = z_i$ este zero. Dacă ipoteza nulă este că mediana diferenței este M_e, atunci putem defini o nouă variabilă $w_i = z_i - M_e$ și aplicăm testul semnelor pentru w_i.

§ 3. Testul medianei

Testul semnelor prezentat în paragraful anterior este aplicabil în cazul în care considerăm selecții de volume egale din două populații. Faptul că cele două selecții erau de același volum făcea ca observațiile să fie considerate perechi.

Dacă datele se referă la fenomene ale căror modalități cantitative sau calitative sînt ordonate după vreun criteriu oarecare, scopul poate fi atins analizînd valorile observate prin testul medianei.

Ipoteza care stă la baza acestui test este următoarea: dacă două selecții bernouliene independente, de n_1 și respectiv n_2 unități, provin dintr-o populație caracterizată de aceeași variabilă aleatoare X, $\frac{n_1}{2}$ și $\frac{n_2}{2}$ dintre unitățile fiecărei selecții ar

trebui să aibă modalităţi $x_i(y_i)$ mai mici decît mediana variabilei aleatoare X, $\frac{n_1}{2}$ şi $\frac{n_2}{2}$ dintre unităţile celor două selecţii ar trebui să aibă modalităţi $x_i(y_i)$ mai mari decît mediana.

Dacă notăm cu $a_1(a_2)$ numărul valorilor $x_i(y_i)$ din prima (a doua) selecţie mai mici decît mediana M_e corespunzătoare celor $n_1 + n_2$ observaţii şi cu $b_1(b_2)$ numărul valorilor $x_i(y_i)$ mai mari decît M_e, repartiţia observată dă naştere la tabelul 5.1.

Tabelul 5.1

	Valori $< M_e$	Valori $> M_e$	Total
Selecţia I	a_1	b_1	$a_1 + b_1$
Selecţia II	a_2	b_2	$a_2 + b_2$
Total	$a_1 + a_2$	$b_1 + b_2$	n

Pentru a determina dacă există sau nu concordanţă între rezultatele observate şi ipoteza H_0 folosim testul lui Fischer (2.8) sau aproximaţia normală (2.7).

§ 4. Testul iteraţiilor

Într-un şir arbitrar cu un număr fixat de elemente, unde fiecare element poate fi de un fel sau mai multe feluri, se numeşte *iteraţie* orice şir de elemente de acelaşi fel. Numărul de elemente ce compun o iteraţie poartă numele de *lungimea iteraţiei*. Numărul total de iteraţii din şir va fi notat cu a. Să presupunem că:

(i) iteraţia conţine numai elemente de tipul a_1 şi a_2;

(ii) u este numărul total de iteraţii din şirul format cu elemente a_1 şi a_2 într-o permutare aleatoare a $n_1 + n_2$ litere dintre care n_1 sînt egale cu a_1 şi n_2 cu a_2;

(iii) toate permutările sînt egal probabile.

Adeseori în locul numărului total de iteraţii u se foloseşte numărul de treceri t.

Scanned with OKEN Scanner

O trecere este definită ca punctul în care se termină o iterație și începe alta.

Numărul de treceri trebuie să fie mai mic cu o unitate decît numărul total de iterații, deoarece ultima iterație, odată închisă, sfîrșește șirul și acest punct nu este numărat ca o trecere. Prin urmare

$$t = u - 1. \tag{4.1}$$

Dacă considerăm drept ipoteză nulă presupunerea că toate permutările sînt egal probabile, atunci

$$M(t) = M(u) - 1, \qquad D^2(t) = D^2(u). \tag{4.2}$$

Se observă că numărul de goluri dintre cele $n_1 + n_2$ numere în care se pot realiza treceri este $n_1 + n_2 - 1$. Fie t_i numărul de treceri în golul i sau cu alte cuvinte

$$t_i = \begin{cases} 0 \text{ dacă } a_1 \text{ și } a_1 \text{ sau } a_2 \text{ și } a_2 \text{ sînt deoparte și alta a golului,} \\ 1 \text{ dacă } a_1 \text{ și } a_2 \text{ sau } a_2 \text{ și } a_1 \text{ sînt deoparte și alta a golului.} \end{cases}$$

În aceste condiții numărul total de treceri este

$$t = \sum_{i=1}^{n_1+n_2-1} t_i \tag{4.3}$$

Avem

$$P(\textit{primul element într-o pereche este } a_1) = \frac{n_1}{n_1 + n_2},$$

$$P(\textit{al doilea element într-o pereche este } a_2 \mid \textit{primul a fost } a_1) = \frac{n_2}{n_1 + n_2 - 1},$$

$$P(\textit{al doilea element într-o pereche este } a_1 \mid \textit{primul a fost } a_1) = \frac{n_1 - 1}{n_1 + n_2 - 1}.$$

de unde

$$P(\textit{ca o pereche să fie formată din } a_1 \textit{ și } a_2) = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 1)},$$

$$P(\textit{ca o pereche să fie formată din } a_1 \textit{ și } a_1) = \frac{n_1(n_1 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)},$$

ceea ce ne permite să întocmim tabelul 5.2.

Scanned with OKEN Scanner

Tabelul 5.2

Felul perechii	t_i	t_i^2	Probabilitate
a_1a_1	0	0	$n_1(n_1-1)/(n_1+n_2)(n_1+n_2-1)$
a_1a_2	1	1	$n_1n_2/(n_1+n_2)(n_1+n_2-1)$
a_2a_1	1	1	$n_1n_2/(n_1+n_2)(n_1+n_2-1)$
a_2a_2	0	0	$n_2(n_2-1)/(n_1+n_2)(n_1+n_2-1)$

Urmează că

$$(4.4)\quad M(t_i) = 1 \cdot \frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)} + \frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)} \cdot 1 = \frac{2n_1n_2}{(n_1+n_2)(n_1+n_2-1)},$$

$$(4.5)\quad M(t_i^2) = 1^2 \cdot \frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)} + \frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)} 1^2 = \frac{2n_1n_2}{(n_1+n_2)(n_1+n_2-1)},$$

de unde

$$M(t) = M\left(\sum_{i=1}^{n_1+n_2-1} t_i\right) = \sum_{i=1}^{n_1+n_2-1} M(t_i) =$$

$$= (n_1+n_2-1)\frac{2n_1n_2}{(n_1+n_2)(n_1+n_2-1)} = \frac{2n_1n_2}{n_1+n_2}$$

și ținînd seama de (4.2)

$$(4.6)\quad M(u) = 1 + \frac{2n_1n_2}{n_1+n_2}.$$

Avem

$$(4.7)\quad D^2(u) = D^2(t) = D^2\left(\sum_{i=1}^{n_1+n_2-1} t_i\right) = \sum_{i=1}^{n_1+n_2-1} D^2(t_i) + 2\sum_{i=1}^{n_1+n_2-2}\sum_{j=i+1}^{n_1+n_2-1} \operatorname{cov}(t_i, t_j)$$

Scanned with OKEN Scanner

și cum

$$\sum_{i=1}^{n_1+n_2-1} D^2(t_i) = (n_1 + n_2 - 1)\, D^2(t_i) = (n_1 + n_2 -$$

$$- 1)\,[M(t_i^2) - M^2(t_i)] = (n_1 + n_2 -$$

$$- 1)\left[\frac{2n_1n_2}{(n_1 + n_2)\,(n_1 + n_2 - 1)} - \left(\frac{2n_1n_2}{(n_1 + n_2)\,(n_1 + n_2 - 1)}\right)^2\right] = \tag{4.8}$$

$$= \frac{2n_1n_2}{n_1 + n_2}\left[1 - \frac{2n_1n_2}{(n_1 + n_2)\,(n_1 + n_2 - 1)}\right] =$$

$$= 2n_1n_2\,\frac{n_1(n_1 - 1) + n_2(n_2 - 1)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

urmează că mai trebuie să determinăm cov $(t_i,\ t_j)$.

Prin definiție

$$\text{cov}\ (t_i,\ t_j) = M(t_it_j) - M(t_i)\,M(t_j)$$

și cum $M(t_i) = M(t_j)$.

$$\text{cov}\ (t_i,\ t_j) = M(t_it_j) - M^2(t_i). \tag{4.9}$$

Termenii t_it_j îi putem considera în două grupe, cu următoarele caracteristici;

1. Termenii t_jt_{j+1} care comportă goluri alăturate, în care perechea de elemente ce determină pe t_j are ca al doilea element primul element al perechii ce determină pe t_{j+1}.

Numărul total de perechi $t_i,\ t_j$ cu $i < j$ este dat de numărul combinărilor ce se pot face cu $n_1 + n_2 - 1$ elemente luate cîte două, adică

$$C^2_{n_1+n_2-1} = \frac{(n_1 + n_2 - 1)\,(n_1 + n_2 - 2)}{2}.$$

Numărul de perechi din această grupă este $n_1 + n_2 - 2$, deoarece, dacă considerăm șirul particular de elemente

$$x_1,\ x_2,\ x_3,\ x_4,\ x_5$$

atunci singurele perechi alăturate de t care pot fi formate sînt $t_1,\ t_2$, din elementele $x_1,\ x_2,\ x_3$; $t_2,\ t_3$ din elementele $x_2,\ x_3,\ x_4$; $t_3,\ t_4$ din elementele $x_3,\ x_4,\ x_5$.

Scanned with OKEN Scanner

Deoarece perechile t_i, t_{i+1} sînt acele perechi de trecere cu un element în comun, toate cazurile posibile sînt trecute în tabelul 5.3.

Tabelul 5.3

Succesiunea de elemente	t_i	t_{i+1}	$t_i t_{i+1}$	Probabilităţile corespunzătoare
$a_1a_1a_1$	0	0	0	0
$a_1a_1a_2$	0	1	0	0
$a_1a_2a_1$	1	1	1	$\frac{n_1}{n_1+n_2}\cdot\frac{n_2}{n_1+n_2-1}\cdot\frac{n_1-1}{n_1+n_2-2}$
$a_1a_2a_2$	1	0	0	0
$a_2a_1a_1$	1	0	0	0
$a_2a_1a_2$	1	1	1	$\frac{n_2}{n_1+n_2}\cdot\frac{n_1}{n_1+n_2-1}\cdot\frac{n_2-1}{n_1+n_2-2}$
$a_2a_2a_1$	0	1	0	0
$a_2a_2a_2$	0	0	0	0

Astfel

$$M(t_it_{i+1}) = 1\cdot\frac{n_1n_2(n_1-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)} + \tag{4.10}$$

$$+1\cdot\frac{n_1n_2(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)} =$$

$$=\frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)},$$

de unde

$$\operatorname{cov}(t_i, t_{i+1}) = \frac{n_1n_2}{(n_1+n_2)(n_1+n_2-1)} - \tag{4.11}$$

$$-\left(\frac{2n_1n_2}{(n_1+n_2)(n_1+n_2-1)}\right)^2$$

şi avem n_1+n_2-2 astfel de termeni.

2. Termenii t_{ij}, unde $j > i+1$, adică golurile de ordin i şi j sînt separate de unul sau mai multe goluri sau cu alte cuvinte

Scanned with OKEN Scanner

termenii pentru care cele două perechi de elemente nu au un element în comun.

Numărul de perechi din această grupă este dat de diferența dintre numărul total de perechi și numărul de perechi din prima grupă, adică de

$$\frac{(n_1+n_2-1)(n_1+n_2-2)}{2}-(n_1+n_2-2)=$$
$$=\frac{(n_1+n_2-2)(n_1+n_2-3)}{2}.$$

Pentru a determina cov (t_i, t_k) cu $k>i+1$ va trebui să determinăm $M(t_it_k)$ pentru datele din tabelul 5.4.

Tabelul 5.4

Tipul perechii	t_i	t_k	t_it_k	Probabilitatea
$a_1a_1 \dots a_1a_1$	0	0	0	
$a_1a_1 \dots a_1a_2$	0	1	0	
$a_1a_1 \dots a_2a_1$	0	1	0	
$a_1a_1 \dots a_2a_2$	0	0	0	
$a_1a_2 \dots a_1a_1$	1	0	0	
$a_1a_2 \dots a_1a_2$	1	1	1	$\frac{n_1n_2(n_1-1)(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)(n_1+n_2-3)}$
$a_1a_2 \dots a_2a_1$	1	1	1	$\frac{n_1n_2(n_1-1)(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)(n_1+n_2-3)}$
$a_1a_2 \dots a_2a_2$	1	0	0	
$a_2a_1 \dots a_1a_1$	1	0	0	
$a_2a_1 \dots a_2a_1$	1	1	1	$\frac{n_1n_2(n_1-1)(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)(n_1+n_2-3)}$
$a_2a_1 \dots a_1a_2$	1	1	1	$\frac{n_1n_2(n_1-1)(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)(n_1+n_2-3)}$
$a_2a_1 \dots a_2a_2$	1	0	0	
$a_2a_2 \dots a_1a_1$	0	0	0	
$a_2a_2 \dots a_1a_2$	0	1	0	
$a_2a_2 \dots a_2a_1$	0	1	0	
$a_2a_2 \dots a_2a_2$	0	0	0	

Scanned with OKEN Scanner

Urmează că

$$\operatorname{cov}(t_i, t_k) = \frac{4n_1n_2(n_1-1)(n_2-1)}{(n_1+n_2)(n_1+n_2-1)(n_1+n_2-2)(n_1+n_2-3)} - \tag{4.12}$$
$$-\left(\frac{2n_1n_2}{(n_1+n_2)(n_1+n_2-1)}\right)^2$$

și avem $(n_1+n_2-2)(n_1+n_2-3)/2$ astfel de termeni.

Ținînd seama în (4.7) de (4.8), (4.11) și (4.12) obținem

$$D^2(u) = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}. \tag{4.13}$$

Dacă selecțiile sînt de volume mari, statistica

$$\frac{u-M(u)}{D(u)} \sim N(0, 1) \tag{4.14}$$

iar regiunea critică se obține pentru acele valori ale statisticii (4.14) care sînt mai mici decît u_α.

Dacă dorim să verificăm ipoteza conform căreia două selecții $(x_1, x_2,\ldots,x_{n_1})$ și $(y_1, y_2,\ldots,y_{n_2})$ provin dintr-o aceeași populație aranjăm cele n_1+n_2 observații în ordinea mărimii și notăm prin u numărul iterațiilor formate cu x și y în acest șir.

Respingem ipoteza emisă, la pragul de semnificație α dacă $u < c$, unde c satisface egalitatea

$$P(u < c) = \alpha.$$

Rezultatele din acest paragraf reprezintă o sistematizare a paragrafului 6.3 din cartea lui K.A. Brownlee [1].

§ 5. Testul Wilcoxon

a) Cu ajutorul acestui test se verifică ipoteza nulă, conform căreia selecțiile $(x_1, \ldots,x_{n_1})$ $(y_1,\ldots y_{n_2})$ provin din populații identice, față de ipoteza alternativă conform căreia populațiile diferă printr-o translație. Vom aranja observațiile din cele două selecții în ordinea mărimii. Se presupune că toate combinațiile rangurilor din aceste două selecții sînt egal probabile. Numărul total de moduri de grupare al rangurilor, dat n_1+n_2 și n_1 este $C^{n_1}_{n_1+n_2}$.

Scanned with OKEN Scanner

Fie R suma rangurilor și $\bar{R}$ media rangurilor unei selecții de volum n_1.

$$(5.1)\quad M(\bar{R}) = \frac{1}{C_{n_1+n_2}^{n_1}}\left[\frac{1}{n_1} C_{n_1+n_2-1}^{n_1-1} \sum_{1}^{n_1+n_2} x_i\right] = \frac{1}{n_1+n_2} \sum_{i=1}^{n_1+n_2} x_i$$

unde x_i sînt întregii de la 1 la $n_1 + n_2$.

Urmează că

$$M(\bar{R}) = \frac{1}{n_1+n_2} \cdot \frac{(n_1+n_2)(n_1+n_2+1)}{2} =$$

$$(5.2)\qquad = \frac{n_1+n_2+1}{2}$$

De asemenea

$$D^2(\bar{R}) = \frac{\sigma^2}{n_1} \frac{n_2}{n_1+n_2-1}$$

și cum

$$\sigma^2 = \frac{1}{n_1+n_2} \frac{(n_1+n_2)(n_1+n_2+1)(2n_1+2n_2+1}{6} -$$

$$(5.3)\qquad -\frac{1}{n_1+n_2}\left(\frac{(n_1+n_2)(n_1+n_2+1)}{2}\right)^2 = \frac{(n_1+n_2)^2-1}{12}$$

obținem

$$(5.4)\quad D^2(\bar{R}) = \frac{(n_1+n_2)^2-1}{12n_1} \cdot \frac{n_2}{n_1+n_2-1} = \frac{(n_1+n_2+1)n_2}{12n_1}.$$

Variabila aleatoare

$$(5.5)\qquad \frac{\bar{R}-M(\bar{R})}{D(\bar{R})} = \frac{\bar{R} - \dfrac{n_1+n_2+1}{2}}{\sqrt{\dfrac{(n_1+n_2+1)n_2}{12n_1}}}$$

este repartizată $N(0,1)$.

b) O generalizare a testului prezentat la punctul a) este dată de Kruskal și Wallis, generalizarea constînd în aceea că în loc de două selecții se consideră k selecții ($k > 2$). Vom aranja ob-

Scanned with OKEN Scanner

servaţiile ce compun cele k selecţii în ordinea mărimii şi vom calcula mediile rangurilor $\bar{R}_i$ pentru fiecare selecţie. Avem

$$M(\bar{R}_i) = \frac{n+1}{2}, \quad n = \sum_{i=1}^{k} n_i$$

şi

$$D^2(\bar{R}_i) = \frac{(n+1)(n-n_i)}{12n_i}$$

ceea ce ne permite să afirmăm că variabila aleatoare

$$\frac{\bar{R}_i - M(\bar{R}_i)}{\sqrt{D^2(\bar{R}_i)}} \sim N(0,1). \tag{5.6}$$

Kruskal şi Wallis au arătat că

$$H = \sum_{i=1}^{k} \frac{\left(\bar{R}_i - \frac{n+1}{2}\right)^2}{\frac{(n+1)(n-n_i)}{12n_i}} \left(1 - \frac{n_i}{n}\right) \sim \chi^2_{(k-1)}. \tag{5.7}$$

(5.7) se mai poate pune sub formele

$$H \simeq \frac{n-1}{n} \sum_{i=1}^{k} \frac{\left(\bar{R}_i - \frac{n+1}{2}\right)^2}{\frac{n^2-1}{12n_i}}, \tag{5.7'}$$

$$H \simeq \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\bar{R}_i^2}{n_i^2} - 3(n+1). \tag{5.7''}$$

Se observă că pentru $k = 2$ statistica H este identică cu pătratul statisticii Wilcoxon (5.5).

c) În cazul în care avem de-a face cu observaţii perechi, se consideră diferenţa dintre prima şi a doua componentă, în valoare absolută, diferenţele rezultate ordonîndu-se după mărime. Rangurile se atribuie începînd cu cea mai mică diferenţă însă diferenţelor negative li se atribuie rang negativ. De exemplu diferenţele

	5,2;	1,3;	1,1;	— 0,2;	0,3;	1,9;	2,7;	— 3,5;	4,7;	5,3;	8
Rangurile	9	4	3	— 1	2	5	6	—7	8	10	11

Scanned with OKEN Scanner

Statistica care se folosește pentru verificarea ipotezei nule conform căreia repartiția diferențelor este simetrică față de zero este dată de suma rangurilor, fiecare rang cu semnul său. Ipoteza pe care o verificăm permite să afirmăm că rangurile au aceeași șansă de a fi pozitive sau negative.

P-cuantila este probabilitatea de a obține o sumă a rangurilor cel mult egală cu valoarea observată.

Deoarece adeseori enumerarea necesară este greoaie se folosește următoarea aproximație normală. Să presupunem că rangurile sînt

$$1,\ 2,\ \ldots\ i,\ \ldots,\ N$$

și să atașăm acestor ranguri variabilele

$$d_1,\ d_2,\ \ldots,\ d_i,\ \ldots,\ d_N.$$

Variabila d_i este atașată rangului i și ia valoarea 0 cînd rangul este negativ și 1 cînd rangul este pozitiv. În ipoteza nulă aceste valori sînt luate cu probabilitatea 1/2. Din definiția variabilelor d_i urmează că suma rangurilor pozitive este

$$S = \sum_{i=1}^{N} i d_i.$$

Deoarece variabila d_i are repartiția

$$d_i : \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

urmează că valoarea medie a variabilei d_i este

$$M(d_i) = \frac{1}{2},$$

de unde

$$M(S) = \sum_{i=1}^{N} iM(d_i) = \frac{1}{2}\sum_{i=1}^{N} i = \frac{1}{2} \cdot \frac{N(N+1)}{2}. \tag{5.8}$$

Avem

$$M(S^2) = M\left(\sum_{i=1}^{N} i d_i\right)^2 = M\left(\sum_{i=1}^{N} i^2 d_i^2 + \sum_{i \neq j}^{N} ij d_i d_j\right)$$

Scanned with OKEN Scanner

și cum

i) $$d_i^2: \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \to M\left(\sum_{i=1}^{N} i^2 d_i^2\right) = \frac{1}{2}\sum_{i=1}^{N} i^2,$$

ii) $$d_i d_j: \begin{pmatrix} 1 & 0 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \to M\left(\sum_{i\neq j} ij d_i d_j\right) = \frac{1}{4}\sum_{i\neq j} ij,$$

obținem

$$M(S^2) = \frac{1}{2}\sum_{i=1}^{N} i^2 + \frac{1}{4}\sum_{i\neq j} ij =$$

$$= \frac{1}{2}\sum_{i=1}^{N} i^2 + \frac{1}{4}\left[\left(\sum_{i=1}^{N} i\right)^2 - \sum_{i=1}^{N} i^2\right] =$$

$$= \frac{1}{4}\sum_{i=1}^{N} i^2 + \frac{1}{4}\left(\sum_{i=1}^{N} i\right)^2. \tag{5.9}$$

Prin urmare dispersia variabilei S este

$$D^2(S) = M(S)^2 - M^2(S) = \frac{1}{4}\sum_{i=1}^{N} i^2 + \frac{1}{4}\left(\sum_{i=1}^{N} i\right)^2 -$$

$$- \left(\frac{1}{2}\sum_{i=1}^{N} i\right)^2 = \frac{1}{4}\sum_{i=1}^{N} i^2 = \frac{1}{4}\cdot\frac{N(N+1)(2N+1)}{6}. \tag{5.10}$$

Variabila aleatoare

$$\frac{S - M(S)}{D(S)}$$

este repartizată $N(0,1)$.

§ 6. Testul X (van der Waerden)

Pentru verificarea ipotezei statistice, asupra provenienței dintr-o aceeași populație a celor două selecții, se poate folosi un test unilateral al lui van der Waerden și anume: se calculează statistica

$$X = \sum_{i=1}^{n_1} \psi\left(\frac{n_i}{n+1}\right) \quad \text{unde} \quad n = n_1 + n_2$$

Scanned with OKEN Scanner

iar ψ este funcția inversă a funcției de repartiție a repartiției normale $N(0,1)$.

Testul poate fi enunțat astfel:

1) ipoteza H se respinge dacă $X > X_\varepsilon$ unde X_ε se determină din relația $P(X > X_\varepsilon) \leqslant \varepsilon$,

2) ipoteza H se acceptă în caz contrar.

Din tabele se extrage X_ε pentru $\varepsilon = 0{,}025,\ 0{,}010,\ 0{,}005$ și pentru un volum de măsurători $n = 6,...,50$.

Exemplul 1: (Al. C. Dorin și V. Gh. Vodă: *Compararea durabilității sculelor prin metode statistice*, „Revista de statistică" nr. 11/1971).

Se compară durabilitățile a două tipuri de burghie elicoidale cu diametrul de 6 mm, unul cu miez îngroșat și altul cu miez normal.

Fie X — variabila aleatoare (caracteristica durabilitate) pentru burghiul cu miez îngroșat, iar Y — caracteristica durabilitate pentru burghiul cu miez normal (tabelul 5.5)

S-a propus să se verifice ipoteza statistică H: variabilele X și Y au aceeași funcție de repartiție față de alternativa H': variabilele X și Y au funcții de repartiție distincte.

Tabelul 5.5

Durabilitatea sculei de tipul I (în minute) x_i	Index r_i	Durabilitatea sculei de tipul I (în minute) x_i	Index r_i	Durabilitatea sculei de tipul II (în minute) y_i	Index r_i	Durabilitatea sculei de tipul II (în minute) y_i	Index r_i
16,10	10	26,50	23	5,05	1	16,10	9
19,18	14	27,22	24	6,50	2	16,30	11
12,25	15	30,26	25	7,16	3	17,54	12
22,58	17	32,18	27	7,20	4	18,02	13
23,20	19	33,58	28	9,40	5	21,28	16
24,44	20	36,58	29	11,34	6	23,06	18
24,58	22	39,56	30	12,18	7	24,58	21
				14,26	8	30,44	26

Pentru calculul mediei și dispersiei fiecărui grup de date alcătuim tabelele 5.6 și 5.7.

Scanned with OKEN Scanner

Tabelul 5.6

nr.	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	16,10	—10,55	111,3025
2	19,18	—7,47	55,8009
3	19,25	—7,40	54,7600
4	22,58	—4,07	16,5649
5	23,20	—3,45	11,9025
6	22,44	—4,21	17,7241
7	24,58	—2,07	4,2849
8	26,50	—0,15	0,0225
9	27,22	0,57	0,3249
10	30,26	3,61	13,0321
11	32,18	5,53	30,5809
12	33,58	6,93	48,0249
13	36,58	9,93	98,6049
14	39,56	12,91	166,6681
	$\sum x_i = 373,21$	—	$\sum (x_i - \bar{x})^2 = 629,5981$

Din tabelul 5.6. rezultă: $\bar{x} = 26,65$, $\sigma_x = 6,95$. Cu ajutorul acestor date calculăm coeficientul de variație

$$V_x = \frac{\sigma_x}{\bar{x}} = 0,26.$$

Tabelul 5.7

nr.	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	5,05	— 9,97	89,4009
2	6,50	— 8,52	72,5904
3	7,16	— 7,86	61,7796
4	7,20	— 7,82	61,1524

Scanned with OKEN Scanner

Tabelul 5.7 (continuare)

nr.	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
5	9,40	— 5,62	31,5844
6	11,34	— 3,68	13,5424
7	12,18	— 2,84	8,0656
8	14,26	— 0,76	0,5776
9	16,10	1,08	1,1664
10	16,30	1,28	1,6384
11	17,54	2,52	6,3504
12	18,02	3,00	9,0000
13	21,28	6,26	39,1876
14	23,06	8,04	64,6416
15	24,58	9,56	91,3936
16	30,44	15,42	237,7764
	$\sum y_i = 240{,}41$		$\sum (y_i - \bar{y})^2 = 799{,}8477$

Din tabelul 5.7 obținem $\bar{y} = 15{,}02$, $\sigma_y = 7{,}30$,

iar coeficientul de variație $V_y = \frac{\sigma_y}{\bar{y}} = 0{,}48$.

Din compararea coeficienților de variație $V_x < V_y$, precum și din însăși valorile lor, rezultă că tendința de grupare a datelor pentru scula de tipul I este mai mare decît la scula de tipul II.

Avem, așadar, o indicație că durabilitățile medii ale celor două tipuri de scule vor fi diferite.

Problema se rezolvă cu ajutorul testului van der Waerden.

În cazul de față avem $n = n_1 + n_2 = 30$, $n + 1 = 31$ iar $n_1 - n_2 = 2$.

Scanned with OKEN Scanner

Se obţine tabelul 5.8 pentru calculul statisticii X.

Tabelul 5.8

$r_i/n+1$	$\psi(r_i/n+1)$	$r_i/n+1$	$\psi(r_i/n+1)$
$\frac{10}{31}=0{,}32$	$-0{,}46$	$\frac{23}{31}=0{,}74$	$+0{,}65$
$\frac{14}{31}=0{,}45$	$-0{,}12$	$\frac{24}{31}=0{,}77$	$+0{,}74$
$\frac{15}{31}=0{,}48$	$-0{,}05$	$\frac{25}{31}=0{,}81$	$+0{,}88$
$\frac{17}{31}=0{,}55$	$+0{,}13$	$\frac{27}{31}=0{,}90$	$+1{,}29$
$\frac{19}{31}=0{,}61$	$+0{,}28$	$\frac{28}{31}=0{,}90$	$+1{,}29$
$\frac{20}{31}=0{,}64$	$+0{,}36$	$\frac{29}{31}=0{,}93$	$+1{,}48$
$\frac{22}{31}=0{,}71$	$+0{,}56$	$\frac{30}{31}=0{,}97$	$+1{,}90$
			$\sum(\cdot)=8{,}93$

Din tabelul 5.8 rezultă tabelul 5.9.

Tabelul 5.9

ε	X_ε
0,025	4,87
0,010	5,75
0,005	6,34

Deci ipoteza se respinge pentru orice risc ales.

Scanned with OKEN Scanner

§ 7. Testul lui Kolmogorov

7.1. Funcția de repartiție empirică

Dacă $x_1, x_2, \ldots, x_n$ este o selecție de volum n și dacă formăm șirul statisticilor de ordine

$$x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$$

atunci funcția $F_n(x)$ definită prin

$$F(x_{(i)}) = P\{X \leqslant x_{(i)}\} = \begin{cases} 0 \text{ dacă } i < 0 \\ \frac{i}{n} \text{ dacă } 1 \leqslant i \leqslant n-1 \\ 1 \text{ dacă } i \geqslant n \end{cases} \tag{7.1}$$

poartă numele de *funcție de repartiție empirică.*

Următoarea teoremă (Glivenko 1933) dă justificarea teoretică a folosirii metodei selecției.

Teorema 1. Notînd prin $F(x)$ funcția de repartiție teoretică și prin $F_n(X)$ funcția de repartiție empirică obținută dintr-o selecție de volum n, avem

$$P(\lim_{n\to\infty} \sup_{x\in R} |F_n(x) - F(x)| = 0) = 1. \tag{7.2}$$

Teorema 2. Dacă $F(x)$ este o funcție continuă, atunci

$$\lim_{n\to\infty} P\left[\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \leqslant \frac{\lambda}{\sqrt{n}}\right] = K(\lambda) = \tag{7.3}$$

$$= \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}$$

pentru $\lambda > 0$.

Pentru demonstrarea acestor teoreme a se vedea Mihoc Gh., Ciucu G., Craiu V. [25].

7.2. Testul lui Kolmogorov

Teorema 2 servește ca bază pentru verificarea concordanței (corespondenței) dintre repartiția teoretică și cea empirică.

Dacă pe baza analizei repartiției empirice se ajunge la concluzia că această repartiție provine dintr-o populație caracteri-

Scanned with OKEN Scanner

zată de o variabilă aleatoare X, avînd funcția de repartiție (teoretică) $F(x)$, și dacă această ipoteză este adevărată, atunci frecvențele relative cumulate au valori apropiate de valorile funcției teoretice de repartiție în punctele respective.

Dacă punem

$$d_n = \max |F_n(x) - F(x)|$$

atunci valoarea variabilei λ se alege astfel încît probabilitatea evenimentului

$$d_n > \frac{\lambda}{\sqrt{n}}$$

să fie mai mică decît pragul de semnificație (de exemplu $\alpha = 0{,}05$) pentru a putea considera că este imposibil să se obțină un astfel de rezultat atunci cînd între funcția de repartiție teoretică și cea empirică există corespondență (atunci cînd ipoteza este adevărată).

Deci probabilitatea de mai înainte este tocmai probabilitatea erorii de ordinul I. Urmează că

$$\alpha = P\left(d_n > \frac{\lambda}{\sqrt{n}}\right).$$

Deoarece evenimentele $d_n > \frac{\lambda}{\sqrt{n}}$ și $d_n \leqslant \frac{\lambda}{\sqrt{n}}$ sînt contrare, urmează că

$$\alpha = 1 - P\left(d_n \leqslant \frac{\lambda}{\sqrt{n}}\right) = 1 - K(\lambda).$$

În tabelul 4 se dau valorile lui $K(\lambda)$ și pentru $\alpha = 0{,}05;\ 0{,}01$ valorile lui $\frac{\lambda}{\sqrt{n}}$ pentru care

$$P\left(d_n \leqslant \frac{\lambda}{\sqrt{n}}\right) = 1 - \alpha. \tag{7.4}$$

În concluzie, dacă $d_n < \frac{\lambda}{\sqrt{n}}$ se acceptă ipoteza conform căreia între funcțiile de repartiție teoretice și empirice există concordanță. Dacă $d_n \geqslant \frac{\lambda}{\sqrt{n}}$, ipoteza se respinge sau cu alte cuvinte

Scanned with OKEN Scanner

regiunea critică pentru verificarea ipotezei este

$$\left\{(x_1, x_2,\ldots,x_n), \max_{-\infty<x<\infty} | F_n(x) - F(x) | > \frac{\lambda}{\sqrt{n}}\right\}$$

unde λ satisface relația (7.4).

Exemplul 2. Să presupunem că aruncînd simultan 4 zaruri s-a observat de cîte ori apare de fiecare dată fețele 4 și 6. S-au obținut rezultatele din tabelul 5.10.

Tabelul 5.10

(1) apariția i	0	1	2	3	4
(2) frecvența observată n_i	11	34	30	6	0
(3) frecvența relativă $\frac{n_i}{n}$	0,1358	0,4198	0,3704	0,0740	0,0000
(4) probabilități $b\left(\cdot, \frac{1}{3}\right)$	0,1975	0,3951	0,2963	0,0987	0,0124
(5) $F(x_i) = \sum_{r=1}^{i} b\left(r, \frac{1}{3}\right)$	0,1975	0,5926	0,8889	0,9876	0,0000
(6) $F_n(x_i)$	0,1358	0,5556	0,9260	1	1
(7) $\lvert F_n(x_i) - F(x_i)\rvert$	0,0617	0,0370	0,0371	0,0124	0

În linia (4) sînt probabilitățile date de binomul $81\left(\frac{2}{3}+\frac{1}{3}\right)^4$. Din tabelul anterior rezultă $d_n = 0{,}0617$, iar din tabela 4, pentru $n = 81$ și $\alpha = 0{,}01$, găsim $\lambda \simeq 1{,}63$, de unde $\lambda/\sqrt{n} = 0{,}181$.

Scanned with OKEN Scanner

Deoarece $d_n = 0{,}0617 < \frac{\lambda}{\sqrt{n}} = 0{,}181$, urmează că repartiția empirică aproximează destul de bine repartiția teoretică.

§ 8. Testul lui Massey

Verificarea normalității pentru eșantioane de volum mic ($10 \leqslant n \leqslant 30$) a făcut obiectul unor articole ale lui F.J. Massey din 1950 și 1951.

Testul lui Massey urmează în linii mari procedura indicată de testul Kolmogorov-Smirnov.

O generalizare a testului lui Massey a fost dată de către Birnbaum și Hall în 1960.

Testul lui Massey constă în următoarele:
fie $x_1, x_2, \ldots, x_n$ selecția de volum n asupra variabilei X pentru care vrem să verificăm ipoteza statistică:

$H_0 : X \sim N(\mu, \sigma^2)$ față de alternativa $H_1 : X \not\in N(\mu, \sigma^2)$.

Se standardizează variabilele $x_{(i)}$

$$y_{(i)} = \frac{x_{(i)} - \bar{x}}{S} \qquad \text{unde } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

iar S este estimatorul abaterii standard teoretice (σ) dat de metoda verosimilității maxime, adică:

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Se calculează $\Phi(y_{(i)}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_{(i)}} \exp\left(-\frac{1}{2} y_{(i)}^2\right) dy_{(i)}$.

Se calculează diferențele $d = |F(y_{(i)}) - \Phi(y_{(i)})|$ și se alege $d_s = \sup d$.

Ipoteza H_0 se acceptă dacă $d_s \leqslant d_g$ unde d_g se extrage din tabelul 5.11 pentru riscul ales și volumul respectiv al selecției.

Scanned with OKEN Scanner

Tabelul 5.11

n	α		n	α	
	0,05	0,10		0,05	0,10
8	0,140	0,163	20	0,117	0,133
9	0,134	0,158	21	0,115	0,131
10	0,130	0,156	22	0,113	0,129
11	0,129	0,155	23	0,112	0,128
12	0,128	0,154	24	0,110	0,126
13	0,128	0,153	25	0,109	0,124
14	0,128	0,151	26	0,108	0,121
15	0,127	0,148	27	0,107	0,120
16	0,126	0.144	28	0,105	0,118
17	0,124	0,142	29	0,104	0,116
18	0,122	0,138	30	0,102	0,114
19	0,120	0,136	32	0,099	0,111

(*Pentru testul Massey;* reprodus după Zbigniew Kowalski [38])

Exemplul 3. (Al. C. Dorin și V. Gh. Vodă: *Compararea durabilității sculelor prin metode statistice*, „Revista de statistică", nr. 11/1971).

Se studiază durabilitatea (în minute) a unui anumit tip de scule aşchietoare (burghiu cu miez normal). Datele de selecție — — scrise în ordinea mărimii sînt: 5, 05, 6,50, 7,16, 7,20, 9,40, 11,34, 12,18, 14,26, 16,10, 16,30, 17,54 18,02, 21,28, 23,06, 24,58, 30,44.

Tabelul 5.12 sintetizează modul de lucru cu testul Massey.

Tabelul 5.12

$x_{(i)}$	$y_{(i)} = \frac{x_{(i)} - \bar{x}}{S}$	Frecvența	Funcția empirică de repartiție	$\Phi(y_{(i)})$	$\lvert F_n(y_{(i)}) - \Phi(y_{(i)}) \rvert$
5,05	—1,36	1	0,06	0,09	0,03
6,50	—1,16	1	0,12	0,13	0,01
7,16	—1,08	1	0,18	0,14	0,04

Scanned with OKEN Scanner

Tabelul 5.12 (continuare)

$x_{(i)}$	$y_{(i)}=\frac{x_{(i)}-\bar{x}}{S}$	Frecvența	Funcția empirică de repartiție	$\Phi(y_{(i)})$	$\lvert F_n(y_{(i)})-\Phi(y_{(i)})\rvert$
7,20	—1,07	1	0,25	0,15	0,10
9,40	—0,77	1	0,31	0,23	0,08
11,34	—0,50	1	0,37	0,31	0,06
12,18	—0,38	1	0,43	0,36	0,07
14,26	—0,10	1	0,50	0,47	0,03
16,10	0,14	1	0,56	0,56	0,00
16,30	0,17	1	0,62	0,57	0,05
17,54	0,34	1	0,68	0,63	0,05
18,02	0,41	1	0,75	0,65	0,10
21,28	0,85	1	0,81	0,80	0,01
26,06	1,10	1	0,87	0,86	0,01
24,58	1,30	1	0,93	0,90	0,03
30,44	2,11	1	1	0,98	0,02

Pentru calcul s-au folosit: $\bar{x} = 15{,}02$ și $S = 7{,}3$.

Se vede că $d_{max} = 0{,}10$. Cum $d_{tabelat} = 0{,}126$ (pentru $n = 16$ și $\alpha = 0{,}05$) se poate trage concluzia că măsurătorile urmează o repartiție aproximativ normală.

§ 9. Testul lui Sarkady

Deoarece în anumite situații testul X^2 nu este semnificativ (micșorarea riscului α ducînd la scăderea puterii testului) pentru verificarea normalității în cazul selecțiilor mici se folosește testul lui Sarkady. Acesta constă în următoarele:

Fie X o variabilă aleatoare a cărei repartiție este necunoscută.

Scanned with OKEN Scanner

Ne propunem să verificăm ipoteza statistică.

$H_0 : X \in N(\mu, \sigma^2)$ față de alternativa $H_1 : X \notin N(\mu, \sigma^2)$. Pentru aceasta se efectuează măsurătorile $x_1,...,x_n$. Se formează variabilele $f_i = x_i - A (i = 1, ..., n-1)$, unde

$$A = \frac{\bar{x}\sqrt{n} + x_1}{\sqrt{n}+1}; \qquad \left(\bar{x} = \frac{1}{n}\sum_1^n x_i\right),$$

și

$$\xi_i = \frac{f_i\sqrt{n-i-1}}{\sqrt{\sum_{i+1}^{n-1} f_k^2}}, \qquad (i = 1, 2,..., n-2).$$

Cu ajutorul tabelului 6 se calculează variabilele

$$\delta_{\nu_i} = S_{\nu_i}(\xi_i) \qquad (i = 1, 2,..., n-2) \tag{9.1}$$

unde S_{ν_i} este funcția de repartiție Student cu $\nu_i = n - i - 1$ grade de libertate, $i = 1, 2,..., n-2$.

Dacă ipoteza H_0 este adevărată, atunci conform unei teoreme din teoria probabilităților, variabilele δ_{ν_i} vor fi repartizate uniform pe segmentul (0,1).

În acest caz s-a redus verificarea ipotezei normalității la verificarea faptului că variabilele δ_{ν_i} posedă repartiția uniformă pe segmentul (0,1).

Pentru verificarea acestei ipoteze este indicat să se aleagă criteriul ω^2 (von Mises-Smirnov) și anume:

Se calculează expresia

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(\eta_i) - \frac{2i-1}{2n}\right]^2 \qquad \text{(M-S)}$$

unde n este volumul eșantionului, iar $F(\eta_i)$ este funcția de repartiție teoretică calculată pentru elementele eșantionului ordonate crescător.

În cazul de față există $n-2$ elemente ale selecției, iar $F(\delta_{(i)}) = \delta_{(i)}$ (proprietatea repartiției uniforme) unde prin $\delta_{(i)}$ s-au notat elementele obținute cu relația (9.1) așezate în ordine crescătoare.

Scanned with OKEN Scanner

Criteriul von Mises-Smirnov capătă forma

$$(n-2)\,\omega_{n-2}^2 = \frac{1}{12(n-2)} + \sum_{i=1}^{n-2}\left[\delta_{(i)} - \frac{2i-1}{2(n-2)}\right]^2;$$

pentru riscul $\alpha = 0{,}05$, decizia de acceptare a ipotezei se ia dacă are loc inegalitatea

$$(n-2)\,\omega_{n-2}^2 \leqslant 0{,}4164.$$

Exemplu. (E. Ursianu și V. Gh. Vodă: *Control statistic de calitate pentru fire relon*, „Materiale Plastice" nr 12/1971).

Se studiază caracteristica „tenacitate" a firului de relon. Pentru verificarea stabilității procesului tehnologic, relativ la această caracteristică, este necesară verificarea normalității acestei caracteristici.

Volumul de date obținut $n = 96$ conduce, dacă aplicăm testul X^2, la respingerea ipotezei de normalitate pentru riscurile uzual alese $\alpha = 0{,}05$ (sau $\alpha = 0{,}025$ și $\alpha = 0{,}01$).

S-a constatat însă că pentru $\alpha = 0{,}0005$ (ceea ce reprezintă un risc foarte mic) ipoteza normalității se acceptă.

În această situație testul X^2 nu este concludent.

Aplicăm testul lui Sarkady.

În cazul caracteristicii „tenacitate", calculele au condus la următoarele rezultate

$$\bar{x} = 4{,}7, \qquad A \simeq 4{,}6, \qquad 94\omega_{94}^2 \simeq 0{,}4500.$$

Așadar, cu un risc $\alpha = 0{,}05$ se poate trage concluzia că variabila „tenacitate" este repartizată normal.

Scanned with OKEN Scanner

CAPITOLUL **6**

ANALIZA DISPERSIONALĂ

§ 1. Analiza dispersională unifactorială

Metoda statistică cunoscută sub numele de analiză dispersională, metodă pe care o vom prezenta în cele ce urmează, a fost introdusă de Fischer. Ea este o metodă de analiză a datelor de observație depinzînd de factori care acționează concomitent, estimînd totodată efectul lor asupra acestora.

Fie $X_1, \ldots, X_k$, k variabile aleatoare independente repartizate $N(\mu_i, \sigma^2)$ unde valorile medii μ_i sînt necunoscute, iar dispersia este aceeași pentru toate variabilele X_i și egală cu σ^2, care este de asemenea necunoscută.

Să presupunem că $x_{i1}, \ldots, x_{in_i}$ sînt n_i, $1 \leqslant i \leqslant k$, observații independente făcute asupra variabilei X_i și că scopul cercetării este să se verifice ipoteza H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, adică ipoteza conform căreia toate variabilele aleatoare considerate au aceeași repartiție. Se observă că avem de-a face cu generalizarea problemei din 4.4.b.

Să introducem următoarele notații:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \tag{1.1}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{1}^{k} n_i \bar{x}_i. \tag{1.2}$$

Deci, $\bar{x}_i$ este media aritmetică a observațiilor din populația caracterizată de variabila aleatoare X_i, iar $\bar{x}$ media tuturor observațiilor, adică media de selecție a întregului spațiu. Numim aceste medii, *media de grup* și respectiv *media totală*.

Scanned with OKEN Scanner

Se observă că

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i+\bar{x}_i-\bar{x})^2=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})^2=$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2+\sum_{i=1}^{k}n_i(\bar{x}_i-\bar{x})^2,$$

deoarece

$$2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(\bar{x}_i-\bar{x})=0.$$

Prin urmare, suma pătratelor abaterilor variabilelor x_{ij} de la media totală $\bar{x}$ poate fi pusă sub forma unei sume de doi termeni. Prima este suma pătratelor abaterilor variabilelor aleatoare de la media lor de grup; numim acest termen *suma abaterilor în interiorul grupurilor*. Al doilea termen este suma abaterilor mediilor de grup de la media totală; numim acest termen *suma abaterilor între grupuri*.

Dacă ipoteza $H_0: \mu_1=\mu_2=\ldots=\mu_k$ este adevărată, toate variabilele aleatoare x_{ij} au aceeași repartiție normală și deoarece ele sînt independente,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=(n-1)\,s^2\sim\chi^2_{n-1}.$$

Din proprietatea de aditivitate a repartiției χ^2, urmează că

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2=\left(\sum_{1}^{k}n_i-k\right)s_2^2\sim\chi^2_{n-k}.$$

Termenul $(n-k)\,s_2^2$ poartă numele de *suma reziduală a pătratelor*.

Dacă ipoteza H_0 este adevărată, expresia

$$\sum_{i=1}^{k}n_i(\bar{x}_i-\bar{x})^2=(k-1)\,s_1^2\sim\chi^2_{k-1}.$$

este independentă de $(n-k)\,s_2^2$.

Scanned with OKEN Scanner

Rezultatele anterioare se pot sintetiza în tabelul 6.1.

Tabelul 6.1

	Suma pătratelor	Numărul gradelor de libertate	Raportul
Între medii	$(k-1)\,s_1^2 = \sum_1^k n_i(\bar{x}_i - \bar{x})^2$	$k-1$	s_1^2
Residual	$(n-k)\,s_2^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	$n-k$	s_2^2
Total	$(n-1)\,s^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$n-1$	s^2

Teorema 1. $s_1^2(s_2^2)\,[s^2]$ *este o estimaţie nedeplasată pentru* σ^2.

Demonstraţie. Avem

$$M(s^2) = \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} M(x_{ij} - \bar{x})^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} M(x_{ij} - \mu + \mu - \bar{x})^2 =$$

$$= \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} \{M[(x_{ij} - \mu)^2] - 2M\,[(x_{ij} - \mu)\,(\mu - \bar{x})] +$$

$$+ M[(\mu - \bar{x})^2]\}.$$

Dacă $\mu_1 = \mu_2 = \ldots = \mu_k$, atunci

$M(\bar{x}) = \mu$,

$$M(\bar{x} - \mu)^2 = D^2(\bar{x}) = M\left(\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}}{n} - \mu \right)^2 =$$

$$= \frac{1}{n^2} M \left[\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \mu) \right]^2$$

Scanned with OKEN Scanner

și cum variabilele aleatoare $(x_{ij} - \mu)$ sînt independente și repartizate $N(0, \sigma^2)$, obținem:

$$M(x_{ij} - \mu)(\bar{x} - \mu) = M(x_{ij} - \mu)\left(\frac{\sum_{\tau=1}^{k}\sum_{l=1}^{n_\tau} x_{\tau l}}{n} - \mu\right) =$$

$$= \frac{1}{n} M(x_{ij} - \mu)\left(\sum_{\tau=1}^{k}\sum_{l=1}^{n_\tau}(x_{\tau l} - \mu)\right) =$$

$$= \frac{1}{n} M\left[\sum_{\tau=1}^{k}\sum_{l=1}^{n_\tau}(x_{ij} - \mu)(x_{\tau l} - \mu)\right] =$$

$$= \frac{1}{n} M\left[(x_{ij} - \mu)^2 + \sum_{\tau,\, l \neq i,\, j}\sum (x_{\tau l} - \mu)(x_{ij} - \mu)\right] = \frac{\sigma^2}{n}.$$

Urmează că

$$M(s^2) = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\sigma^2 - 2\frac{\sigma^2}{n} + \frac{\sigma^2}{n}\right) = \sigma^2.$$

Avem

$$M(s_1^2) = \frac{1}{k-1} M\left[\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2\right] = \frac{1}{k-1} M\left\{\sum_{i=1}^{k} n_i[(\bar{x}_i - \mu) -\right.$$

$$\left. - (\bar{x} - \mu)]^2\right\} = \frac{1}{k-1}\sum_{i=1}^{k} n_i M\left[(\bar{x}_i - \mu)^2 -\right.$$

$$\left. - 2\left(\frac{\bar{x}_i - \mu}{n}\right)\left(\sum_{j=1}^{k} n_j(\bar{x}_j - \mu)\right) + (\bar{x} - \mu)^2\right] =$$

$$= \frac{1}{k-1}\sum_{i=1}^{k} n_i M\left[(\bar{x}_i - \mu)^2 - \frac{2n_i}{n}(\bar{x}_i - \mu)^2 -\right.$$

$$\left. - \frac{2}{n}\sum_{\substack{j=1 \\ j\neq 1}}^{k} n_j(\bar{x}_i - \mu)(\bar{x}_j - \mu) + (\bar{x} - \mu)^2\right] =$$

$$= \frac{1}{k-1}\sum_{i=1}^{k} n_i\left[\frac{\sigma^2}{n_i} - \frac{2n_i}{n}\frac{\sigma^2}{n_i} + \frac{\sigma^2}{n}\right] = \sigma^2.$$

Scanned with OKEN Scanner

Analog se arată că $M(s_2^2) = \sigma^2$, afirmaţia adevărată chiar dacă ipoteza H_0 nu este adevărată.

Deoarece $(n-1)s^2$ a fost despărţită în doi termeni independenţi, se obişnuieşte să se numească această despărţire *analiza dispersiei*. Deoarece $(n-k)s_2^2$ şi $(k-1)s_1^2$ sînt variabile aleatoare independente şi repartizate χ^2 cu $(n-k)$ şi respectiv $(k-1)$ grade de libertate, urmează că variabila aleatoare

$$F = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/(n-k)} \sim F_{k-1,\, n-r}.$$

Respingem ipoteza H_0 dacă valoarea observată a lui F este $\geqslant F_{\alpha;\, k-1,\, n-r}$, determinat din tabelele 5.

§ 2. Analiza dispersională bifactorială şi cu o observaţie în fiecare grup

Să presupunem că observaţiile sînt împărţite în r grupe conform factorului A şi în v grupe conform factorului B. Vom considera cazul în care avem o observaţie în fiecare grupă, adică în total $n = rv$ observaţii.

Fie x_{kl} observaţia care aparţine grupului k conform factorului A şi grupul l conform factorului B.

Această clasificare este reprezentată în tabelul 6.2

Tabelul 6.2

A \ B	1	2	...	v	$\bar{x}_{k.}$
1	x_{11}	x_{12}	...	x_{1v}	$\bar{x}_{1.}$
2	x_{21}	x_{22}	...	x_{2v}	$\bar{x}_{2.}$
⋮	⋮	⋮		⋮	⋮
r	x_{r1}	x_{r2}	...	x_{rv}	$\bar{x}_{r.}$
$\bar{x}_{.l}$	$\bar{x}_{.1}$	$\bar{x}_{.2}$	...	$\bar{x}_{.v}$	$\bar{x}$

Scanned with OKEN Scanner

unde

$$\bar{x}_{k\cdot} = \frac{1}{v}\sum_{l=1}^{v} x_{kl}, \quad \bar{x}_{\cdot l} = \frac{1}{r}\sum_{k=1}^{r} x_{kl}, \quad \bar{x} = \frac{1}{rv}\sum_{k=1}^{r}\sum_{l=1}^{v} x_{kl}.$$

Se observă că

$$\sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x})^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x} + \bar{x}_{k\cdot} -$$

$$- \bar{x} + \bar{x}_{\cdot l} - \bar{x})^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x})^2 +$$

$$+ v\sum_{k=1}^{r}(\bar{x}_{k\cdot} - \bar{x})^2 + r\sum_{l=1}^{v}(\bar{x}_{\cdot l} - \bar{x})^2.$$

Presupunem că variabilele aleatoare $(x_{kl})_{\substack{1\leqslant k\leqslant r\\ 1\leqslant l\leqslant v}}$ sînt repartizate $N(\mu_{kl}, \sigma^2)$ unde σ^2 este necunoscută. Ne propunem să verificăm ipoteza H_0 conform căreia toate valorile medii μ_{kl} sînt egale.

Dacă ipoteza H_0 este adevărată, atunci

$$M(\bar{x}_{\cdot l}) = \frac{1}{r}\sum_{k=1}^{r} M(x_{kl}) = \frac{r\cdot\mu}{r} = \mu,$$

$$M(\bar{x}_{k\cdot}) = \frac{1}{v}\sum_{l=1}^{v} M(x_{kl}) = \frac{v\cdot\mu}{v} = \mu,$$

$$M(\bar{x}) = \frac{1}{rv}\sum_{k=1}^{r}\sum_{l=1}^{v} M(x_{kl}) = \frac{r\cdot v\mu}{rv} = \mu.$$

Urmează că

$$M(x_{kl} - \bar{x}) = M(x_{kl}) - M(\bar{x}) = 0,$$

$$D^2(x_{kl} - \bar{x}) = M(x_{kl} - \bar{x})^2 = M\,[(x_{kl} - \mu) - (\bar{x} - \mu)]^2 =$$

$$= M\,[(x_{kl} - \mu)^2 - 2(x_{kl} - \mu)(\bar{x} - \mu) + (\bar{x} - \mu)^2] =$$

$$= M\left[(x_{kl} - \mu)^2 - 2(x_{kl} - \mu)\left(\frac{\sum_{i=1}^{r}\sum_{j=1}^{v} x_{ij}}{rv} - \mu\right) + (\bar{x} - \mu)^2\right] =$$

$$= M(x_{kl} - \mu)^2 - \frac{2}{rv}M(\Sigma\,(x_{kl} - \mu)(x_{ij} - \mu)) + M(\bar{x} - \mu)^2,$$

Scanned with OKEN Scanner

de unde

$$D^2(x_{kl} - \bar{x}) = \sigma^2 - \frac{2}{rv}\sigma^2 - \frac{1}{rv}\sigma^2 = \sigma^2\left(\frac{rv-1}{rv}\right).$$

Variabilele $x_{kl} - \bar{x}$ sînt independente și repartizate $N\left(0, \sigma^2\frac{rv-1}{rv}\right)$.

Deoarece

$$\sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}) = \sum_{k=1}^{r}\sum_{l=1}^{v}x_{kl} - rv\bar{x} = 0,$$

urmează că variabilele aleatoare $(rv-1)\,s^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x})^2$ este repartizată χ^2 cu $rv - 1$ grade de libertate.

Fie

$$v\sum_{1}^{r}(\bar{x}_{k\cdot} - \bar{x})^2 = (r-1)\,s_1^2, \qquad r\sum_{l=1}^{v}(\bar{x}_{\cdot l} - \bar{x})^2 = (v-1)\,s_2^2.$$

Deoarece $\sum_{1}^{r}(\bar{x}_{k\cdot} - \bar{x}) = \sum_{1}^{r}\bar{x}_{k\cdot} - r\bar{x} = 0$, înseamnă că forma pătratică $(r-1)\,s_1^2$ are rangul $r - 1$.

Deoarece $\sum_{1}^{v}(\bar{x}_{\cdot l} - \bar{x}) = \sum_{1}^{v}\bar{x}_{\cdot l} - v\bar{x} = 0$, urmează că forma pătratică $(v-1)\,s_2^2$ are rangul $v - 1$.

Fie

$$\sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x})^2 = (r-1)(v-1)\,s_3^2.$$

Avem

$$\sum_{k=1}^{r}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x}) = r\bar{x}_{\cdot l} - r\bar{x} - r\bar{x}_{\cdot l} + r\bar{x} = 0,$$

$$1 \leqslant l \leqslant v,$$

$$\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x}) = v\bar{x}_{k\cdot} - v\bar{x}_{k\cdot} - v\bar{x} + v\bar{x} = 0,$$

$$1 \leqslant k \leqslant r$$

Scanned with OKEN Scanner

și cum pentru $k = l$, relațiile

$$\sum_{k=1}^{r} (x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x}) = 0$$

și

$$\sum_{l=1}^{v} (x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x}) = 0$$

coincid, urmează că forma pătratică $(r-1)(v-1)s_3^2$ are $rv - r - v + 1 = (r-1)(v-1)$ grade de libertate.

Dacă ipoteza H_0 este adevărată, și cum $(rv-1)s^2 = \sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x})^2$ este repartizată χ^2 cu $(rv-1)$ grade de libertate, urmează că $(r-1)s_1^2$, $(v-1)s_2^2$ și $(r-1)(v-1)s_3^2$ sînt independente și repartizate χ^2 cu $(r-1)$, $(v-1)$ și respectiv $(r-1)(v-1)$, grade de libertate.

Teorema 2. *s^2, s_1^2, s_2^2, s_3^2 sînt estimații nedeplasate pentru σ^2.*

Demonstrație. Deoarece

$$M(x_{kl} - \bar{x})^2 = \frac{\sigma^2(rv-1)}{rv}, \qquad \begin{matrix} 1 \leqslant k \leqslant r \\ 1 \leqslant l \leqslant v \end{matrix}$$

urmează că

$$M(s^2) = \frac{1}{rv-1}\sum_{k=1}^{r}\sum_{l=1}^{v} M(x_{kl} - \bar{x})^2 = \frac{1}{rv-1}\sum_{k=1}^{r}\sum_{l=1}^{v}\frac{rv-1}{rv}\sigma^2 = \sigma^2.$$

Avem

$$M(\bar{x}_{k\cdot} - \bar{x}) = 0,$$

$$M(\bar{x}_{k\cdot} - \bar{x})^2 = M[(\bar{x}_{k\cdot} - \mu) - (\bar{x} - \mu)]^2 =$$

$$= M\left[(\bar{x}_{k\cdot} - \mu)^2 - 2(\bar{x}_{k\cdot} - \mu)\left(\frac{\sum_{1}^{r}\bar{x}_{i\cdot}}{r} - \mu\right) + (\bar{x} - \mu)^2\right] =$$

$$= M(\bar{x}_{k\cdot} - \mu)^2 - \frac{2}{r}M(\bar{x}_{k\cdot} - \mu)^2 - \frac{2}{r}\sum_{i\neq k}M(\bar{x}_{k\cdot} - \mu)(\bar{x}_{i\cdot} - \mu) +$$

$$+ M(\bar{x} - \mu)^2] = \frac{\sigma^2}{v} - \frac{2\sigma^2}{rv} + \frac{\sigma^2}{rv} = \frac{\sigma^2}{rv}(r-1),$$

Scanned with OKEN Scanner

de unde

$$M(s_1^2) = \frac{v}{r-1}\sum_1^r (\bar{x}_{k\cdot} - \bar{x})^2 = \frac{v}{r-1}\cdot\frac{r-1}{rv}\,r\sigma^2 = \sigma^2.$$

Avem

$$M(\bar{x}_{\cdot l} - \bar{x}) = 0,$$

$$M(\bar{x}_{\cdot l} - \bar{x})^2 = M[(\bar{x}_{\cdot l} - \mu) - (\bar{x} - \mu)]^2 =$$

$$= M\left[(\bar{x}_{\cdot l} - \mu)^2 - 2(\bar{x}_{\cdot l} - \mu)\left(\frac{\sum_1^v \bar{x}_{\cdot j}}{v} - \mu\right) + (\bar{x} - \mu)^2\right] =$$

$$= M(\bar{x}_{\cdot l} - \mu)^2 - \frac{2}{v}M(\bar{x}_{\cdot l} - \mu)^2 - \frac{2}{v}\sum_{j\neq l} M(\bar{x}_{\cdot l} - \mu)(\bar{x}_{\cdot j} - \mu) +$$

$$+ M(\bar{x} - \mu)^2 = \frac{\sigma^2}{r} - \frac{2}{v}\frac{\sigma^2}{r} + \frac{\sigma^2}{rv} = \sigma^2\frac{v-1}{rv},$$

de unde

$$M(s_2^2) = \frac{r}{v-1}\sum_1^v M(\bar{x}_{\cdot l} - \bar{x})^2 = \frac{r}{v-1}\,v\,\frac{v-1}{rv}\,\sigma^2 = \sigma^2.$$

Analog, se arată că

$$M(s_3^2) = \frac{1}{(r-1)(v-1)}\sum_{k=1}^r\sum_{l=1}^v M(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x})^2 = \sigma^2.$$

Rezultatele obținute sînt sintetizate în tabelul 6.3

Tabelul 6.3

	Suma pătratelor	Numărul gradelor de libertate	Raportul
În linii	$(r-1)s_1^2 = v\sum_1^r (\bar{x}_{k\cdot} - \bar{x})^2$	$r-1$	s_1^2
în coloane	$(v-1)s_2^2 = r\sum_1^v (\bar{x}_{\cdot l} - \bar{x})^2$	$v-1$	s_2^2
rezidual	$(r-1)(v-1)s_3^2 = \sum_{k=1}^r\sum_{l=1}^v (x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x})^2$	$(r-1)(v-1)$	s_3^2
Total	$(rv-1)s^2 = \sum_{k=1}^r\sum_{l=1}^v (x_{kl} - \bar{x})^2$	$rv-1$	s^2

Scanned with OKEN Scanner

Fie

$$\bar{\mu}_{k\cdot} = \frac{1}{v}\sum_{l=1}^{v}\mu_{kl}, \qquad \bar{\mu}_{\cdot l} = \frac{1}{r}\sum_{k=1}^{r}\mu_{kl},$$

$$\bar{\mu} = \frac{1}{rv}\sum_{k=1}^{r}\sum_{l=1}^{v}\mu_{kl} = \frac{1}{r}\sum_{k=1}^{r}\bar{\mu}_{k\cdot} = \frac{1}{v}\sum_{l=1}^{v}\bar{\mu}_{\cdot l}$$

și ipotezele

$$H_A : \bar{\mu}_{1\cdot} = \bar{\mu}_{2\cdot} = \ldots = \bar{\mu}_{r\cdot},$$

$$H_B : \bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2} = \ldots = \bar{\mu}_{\cdot v}.$$

Dacă ipoteza H_0 este adevărată, atunci sînt adevărate și ipotezele H_A și H_B.

Pentru a verifica ipoteza H_0, din tabelele 5, găsim $F_{r-1,\,(r-1)(v-1);\alpha}$

Dacă $F_{r-1,\,(r-1)(v-1):\alpha} < F_A$ respingem ipoteza H_A, unde

$$F_A = \frac{s_1^2}{s_3^2} = \frac{(r-1)(v-1)}{(r-1)} \cdot \frac{v\sum_{1}^{r}(\bar{x}_{k\cdot} - \bar{x})^2}{\sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - \bar{x}_{\cdot l} + \bar{x})^2}$$

este repartizat F cu $(r-1)$ și $(r-1)(v-1)$ grade de libertate.

Analog, pentru a verifica ipoteza H_B, din tabelele 5, găsim

$$F_{v-1,\,(r-1)\,v-1);\,\alpha}.$$

Dacă $F_{v-1,\,(r-1)\,(v-1);\,\alpha} < F_B$, respingem ipoteza H_B, unde

$$F_B = \frac{s_2^2}{s_3^2} = \frac{(r-1)(v-1)}{(v-1)} \cdot \frac{r\sum_{1}^{v}(\bar{x}_{\cdot l} - \bar{x})^2}{\sum_{k=1}^{r}\sum_{l=1}^{v}(x_{kl} - \bar{x}_{k\cdot} - x_{\cdot l} + \bar{x})^2}$$

este repartizată F cu $(v-1)$ și $(r-1)(v-1)$ grade de libertate.

Exemplu. (T. Baron: *Programarea experiențelor privind calitatea produselor*, Revista de statistică, 4/1969).

Printre metodele de planificare a experiențelor se numără și metoda blocurilor cu dispunerea întîmplătoare a obiectelor experimentului sau metoda blocurilor randomizate. Această metodă se caracterizează prin aceea că numărul total rk al expe-

Scanned with OKEN Scanner

rienţelor se împarte în r părţi omogene, numite blocuri, fiecare cuprinzînd cîte k obiecte de experienţă. Obiectele de experienţă se dispun în interiorul blocurilor în mod întîmplător între procedeele de prelucrare. Pentru fiecare bloc se efectuează un anumit procedeu de prelucrare.

Se analizează rezultatele unei experienţe la o fabrică de tricotaje unde s-a luat în studiu influenţa a patru procedee tehnologice (P_1, P_2, P_3, P_4), asupra rezistenţei la plesnire a unui tricot.

S-au supus experienţei 16 obiecte care s-au împărţit în 4 blocuri. În urma extracţiei întîmplătoare s-a obţinut următoarea dispunere a procedeelor de prelucrare în cadrul fiecărui bloc:

$$P_3P_4P_1P_2, \quad P_2P_4P_3P_1, \quad P_2P_3P_1P_4, \quad P_1P_4P_3P_2$$

S-au obţinut rezultatele din tabelul 6.4.

Tabelul 6.4

Procedee i	Blocuri				$\sum_j x_{ij}$	$\bar{x}_{i\cdot}$
	1	2	3	4		
P_1	1,46	2,00	1,48	1,51	6,45	1,61
P_2	2,58	3,03	2,82	2,90	11,33	2,83
P_3	4,15	4,61	4,31	4,13	17,20	4,30
P_4	4,54	4,52	4,53	4,15	17,74	4,44
$\sum_i x_{ij}$	12,73	14,16	13,14	12,69	52,72	
$\bar{x}_{\cdot j}$	3,18	3,54	3,29	3,17		

Împrăştierea totală, caracterizată prin suma pătratelor diferenţelor dintre valorile caracteristicii şi media acestor valori, se descompune în: 1) împrăştierea între blocuri, 2) împrăştierea între procedeele de prelucrare şi 3) împrăştierea reziduală care caracterizează variaţia caracteristicii în interiorul blocurilor.

Analiza rezultatelor experienţelor (analiza împrăştierii) o vom face cu ajutorul analizei dispersionale după schema de lucru din tabelul 6.5 analog cu 6.3.

Scanned with OKEN Scanner

Tabelul 6.5

Împrăștierea	Suma pătratelor abaterilor	Nr. gradelor de libertate	Dispersia	Statistica F
Între blocuri	$k\sum_j (\bar{x}_{\cdot j} - \bar{x})^2$	$r - 1$	s_3^2	s_3^2/s_1^2
Între procedeele de prelucrare	$r\sum_i (\bar{x}_{i\cdot} - \bar{x})^2$	$k - 1$	s_2^2	s_2^2/s_1^2
Reziduală	$\sum_i \sum_j (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$	$(r-1)(k-1)$	s_1^2	
Totală	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	$(rk - 1)$		

Pentru obținerea sumelor pătratelor abaterilor, în tabelul 6.6 sînt determinate pătratele valorilor observate x_{ij}^2.

Tabelul 6.6

Procedee	Blocuri : j				$\sum_i x_{ij}^2$	$(\sum_j x_{ij})^2$
	1	2	3	4		
P_1	2,1316	4,0000	2,1904	2,2801	10,6021	41,6025
P_2	6,6564	9,1809	7,9524	8,4100	32,1997	128,3689
$i\ P_3$	17,2225	21,2521	18,5761	17,0569	74,1076	295,8400
P_4	20,6116	20,4304	20,5209	17,2225	78,7854	314,7076
$\sum_i x_{ij}^2$	46,6221	54,8634	49,2398	44,9695	195,6948	780,5190
$\left(\sum_i x_{ij}\right)^2$	162,0529	200,5056	172,65	161,0361	696,2542	2779,3984

$$\frac{(\sum_i \sum_j x_{ij})^2}{rk} = \frac{(52,72)^2}{16} = \frac{2779,3984}{16} = 173,7124.$$

Scanned with OKEN Scanner

Sumele pătratelor abaterilor vor fi

$$k\sum_j (\bar{x}_{\cdot j} - \bar{x})^2 = \sum_j \frac{(\sum_i x_{ij})^2}{k} - \frac{(\sum_i \sum_j x_{ij})^2}{rk} = \frac{696{,}2542}{4} -$$

$$- 173{,}7124 = 0{,}3512, \qquad (1)$$

$$r\sum_i (\bar{x}_{i\cdot} - \bar{x})^2 = \sum_i \frac{(\sum_j x_{ij})^2}{r} - \frac{(\sum_i \sum_j x_{ij})^2}{rk} = \frac{780{,}5192}{4} -$$

$$- 173{,}7124 = 24{,}4174,$$

$$\sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2 = \sum_i \sum_j x_{ij}^2 - \sum_j \frac{(\sum_i x_{ij})^2}{k} -$$

$$- \sum_i \frac{(\sum_j x_{ij})^2}{r} + \frac{(\sum_i \sum_j x_{ij})^2}{rk} = 195{,}6948 - 174{,}06355 - 195{,}1298 +$$

$$+ 173{,}7124 = 0{,}2138, \qquad (3)$$

$$\sum_i \sum_j x_{ij}^2 = (1{,}46)^2 + (2{,}00)^2 + \ldots + (4{,}15)^2 = 195{,}6948,$$

$$\sum_i \sum_j (x_{ij} - \bar{x})^2 = \sum_i \sum_j x_{ij}^2 - \frac{(\sum_i \sum_j x_{ij})^2}{rk} = 195{,}6948 -$$

$$- 173{,}7124 = 21{,}9824 \qquad (4)$$

Rezultatele calculelor le vom introduce în tabelul 6.7.

Tabelul 6.7

Împrăștierea	Suma pătratelor	Nr. gradelor de libertate	Dispersia	*F*
Între blocuri	0,3512	3	0,1171	4,93
Între procedee	21,4174	3	7,1391	300,5
Rămasă	0,2138	9	0,0238	
Totală	21,9824	15		

Scanned with OKEN Scanner

Din tabelul 5a se găseşte $F_{0,95;\,3;\,9} = 3{,}86$. Comparînd această valoare tabelară cu $F = 4{,}93$, constatăm că mediile pe blocuri diferă semnificativ, adică grupurile de obiecte de experienţă nu sînt omogene. Acelaşi rezultat îl obţinem şi pentru verificarea semnificaţiei diferenţelor mediilor pe procedee de prelucrare ($300{,}5 > 3{,}86$) ceea ce ne determină să tragem concluzia că cele patru procedee de prelucrare dau rezultate diferite.

În continuare se verifică care dintre procedee duce la obţinerea de rezultate diferite adică se stabileşte care dintre mediile celor patru procedee de prelucrare se deosebeşte semnificativ de celelalte şi care sînt acele medii care diferă numai întîmplător între ele. Acest lucru se realizează eliminîndu-se treptat din studiu acel procedeu a cărui medie se abate cel mai mult faţă de mediile celorlalte procedee, deci şi faţă de media generală.

Din tabelul 6.4, coloana ultimă, rezultă că trebuie eliminat procedeul P_1 a cărei medie este 1,61. Evident, se pune întrebarea dacă se exclude procedeul P_1, vor exista încă diferenţe semnificative între P_2, P_3 şi P_4? Pentru a răspunde vom relua calculele după metodologia expusă mai înainte, folosind însă datele din tabelul 6.8.

Tabelul 6.8.

	B_1	B_2	B_3	B_4	$\sum_j x_{ij}$	$(\sum_i x_{ij})^2$
P_2	2,58	3,03	2,82	2,90	11,33	128,369
P_3	4,15	4,61	4,31	4,13	17,20	295,840
P_4	4,54	4,52	4,53	4,15	17,74	314,708
$\sum_i x_{ij}$	11,27	12,16	11,66	11,18	46,27	738,917
$\left(\sum_i x_{ij}\right)^2$	127,013	147,866	135,956	124,992	535,827	

$$i = 1, \ldots, k; \quad j = 1, \ldots, r; k = 3; \quad r = 4,$$

$$x_{ij}^2 = 195{,}6948 - 10{,}6021 = 185{,}0927,$$

$$\left(\sum_i \sum_j x_{ij}\right)^2 = (46{,}27)^2 = 2\,140{,}9129.$$

Scanned with OKEN Scanner

Rezultă în continuare valorile împrăștierii din tabelul 6.9.

Tabelul 6.9

Împrăștierea	Suma pătratelor	Nr. gradelor de libertate	Dispersia	F
Între blocuri	0,1994	3	0,0665	2,43
Între procedee	6,3197	2	3,1599	115,59
Rămasă	0,1642	6	0,0274	
Totală	6,6833	11		

$$F_{0,95;\,3;\,6} = 4,76; \quad F_{0,95;\,2;\,6} = 5,14.$$

De data aceasta blocurile nu mai diferă semnificativ între ele ($2,43 < 4,76$), deci au devenit omogene. Procedeele de prelucrare rămase în studiu (P_2, P_3 și P_4) continuă însă să difere semnificativ ($115,59 > 5,14$).

Vom renunța la procedeul P_2. Excluzînd din calcul valorile acestui procedeu obținem

$$\sum_i \sum_j x_{ij} = 46,27 - 11,38 = 34,89;$$

$$\sum_i \sum_j x_{ij}^2 = 185,0927 - 32,1997 = 152,8930; \; k = 2; \; r = 4.$$

Tabelul 6.10

Împrăștierea	Suma pătratelor	Nr. gradelor de libertate	Dispersia	F
Între blocuri	0,1881	3	0,0627	2,76
Între procedee	0,0364	1	0,0364	1,61
Rămasă	0,0680	3	0,0227	
Totală	0,2925	7		

$$F_{3;\,3;\,0,95} = 9,25; \qquad F_{1;\,3;\,0,95} = 10,1$$

Procedeele P_3 și P_4 de prelucrare nu diferă semnificativ. Rezultă, așadar, că pentru a obține aceleași rezultate asupra caracteristicii cercetate va trebui să lucrăm numai cu procedeele P_3 și P_4.

Scanned with OKEN Scanner

CAPITOLUL
7

REGRESIE LINEARĂ

§ 1. Introducere

Considerăm două variabile aleatoare x și y și problema determinării gradului de legătură stochastică dintre ele. Cercetarea formei legăturii dintre două astfel de variabile, bazată pe k perechi de observații

$$(x_1, y_1), \quad (x_2, y_2), \ldots, (x_k, y_k),$$

este ușurată de trecerea acestor perechi de observații într-un sistem rectangular de axe. Graficul astfel obținut, poartă numele de *diagramă de împrăștiere*. Această diagramă arată dacă între cele două variabile există o relație și dacă legătura poate fi considerată lineară. Pentru a măsura gradul în care aceste variabile sînt linear legate, cerem ca indicatorul folosit să satisfacă condițiile:

a) să fie independente față de alegerea originii pentru variabile;

b) să fie independent de unitățile de măsură folosite pentru x și y.

Aceste două condiții sînt îndeplinite dacă indicatorul folosit este construit considerînd variabilele x_i și y_i în forma $u_i = (x_i - \bar{x})/s_x$ și $(y_i - \bar{y})/s_y = v_i$. Dacă considerăm diagrama de împrăștiere a punctelor (u_i, v_i) vom constata că cele mai multe puncte se găsesc în cadranele 1 și 3 și că coordonatele acestor puncte au valori mai mari, decît ale punctelor din cadranele 2 și 4.

O măsură simplă a acestei împrăștieri este dată de suma $\sum_1^k u_i v_i$. Termenii acestei sume corespunzători punctelor din cadra-

Scanned with OKEN Scanner

nele 1 şi 3 vor fi pozitivi, în timp ce aceia din cadranele 2 şi 4 vor fi negativi. Prin urmare o valoare pozitivă mare a acestei sume va indica o puternică tendinţă lineară în diagrama de împrăştiere. Acest lucru nu este întotdeauna adevărat, deoarece, prin dublarea numărului de puncte fără schimbarea naturii împrăştierii, valoarea acestei sume va fi aproximativ dublă. Prin urmare este necesar să se împartă această sumă la numărul punctelor, înainte de a se folosi ca măsură a legăturii dintre x şi y.

Suma $\sum_{1}^{k} u_i v_i / k$, numită *coeficient de corelaţie de selecţie* şi notată prin r, este indicator pentru legătura lineară dintre variabilele x şi y.

Mărimea lui r determină intensitatea acestei legături în timp ce semnul lui r arată tendinţa lui y de a creşte sau descreşte, cu x.

r variază între -1 şi $+1$, ia valoarea ± 1 dacă şi numai dacă, punctele diagramei de împrăştiere sînt situate pe o linie dreaptă.

Vom reveni asupra coeficientului de corelaţie de selecţie în capitolul 8.

Studiul legăturii dintre două variabile poate fi folosit la estimarea uneia dintre variabile. Coeficientul de corelaţie nefiind în măsură să rezolve o astfel de problemă, apelăm la metoda regresiei.

Dacă din diagrama de împrăştiere rezultă că x şi y sînt aproximativ linear legate pentru estimarea lui y pentru x, este suficient să folosim o funcţie lineară de x. Astfel, problema estimării cere în primul rînd să se ajusteze mulţimea de puncte printr-o linie dreaptă.

Dreapta de ajustare este cunoscută sub numele de *dreaptă de regresie.*

§ 2. Modelul

Problema ajustării mulţimii de puncte printr-o linie dreaptă revine la problema estimării parametrilor curbei. Cu toate că există numeroase metode pentru estimare, cea mai bună metodă este metoda celor mai mici pătrate.

Scanned with OKEN Scanner

În cele ce urmează vom presupune că y este repartizat $N(\mu, \sigma^2)$ şi că toate observaţiile sînt independente. De asemenea, vom mai presupune că μ este o funcţie lineară de x, adică

$$\mu = \alpha + \beta(x - \bar{x}), \quad \bar{x} = \frac{1}{k}\sum_1^k x_i. \tag{2.1}$$

Se pune problema să găsim estimaţii pentru α, β şi σ^2 (le notăm prin a, b şi respectiv s^2) şi să determinăm repartiţia lor. Dreapta de regresie estimatoare este

$$Y = a + b(x - \bar{x}), \tag{2.2}$$

unde b este panta, iar a este punctul de intersecţie al dreptei (2.1) cu dreapta $x = \bar{x}$.

Eroarea de estimare este dată de diferenţa dintre valorile observate y_i şi Y-cii obţinuţi prin înlocuirea x-lor în dreapta de regresie (2.2). Deoarece aceste diferenţe pot fi pozitive sau negative, pentru o dreaptă de ajustare cît mai bună, cerem ca suma pătratelor erorilor să fie minimă. Valorile parametrilor obţinuţi prin acest procedeu determină dreapta de ajustare în sensul metodei celor mai mici pătrate.

Astfel facem minimă.

$$S = \sum_1^k (y_i - Y_i)^2 = \sum_1^k [y_i - a - b(x_i - \bar{x})]^2. \tag{2.3}$$

Pentru a determina valorile lui a şi b care fac minimă suma S, diferenţiem (2.3) în raport cu a şi b şi egalăm cu zero. Obţinem

$$\frac{\partial S}{\partial a} = -2\Sigma[y_i - a - b(x_i - \bar{x})] = 0,$$

$$\frac{\partial S}{\partial b} = -2\Sigma[y_i - a - b(x_i - \bar{x})](x_i - \bar{x}) = 0,$$

sau

$$\begin{aligned} ka + b\sum_1^k (x_i - \bar{x}) &= \sum_1^k y_i, \\ a\sum_1^k (x_i - \bar{x}) + b\sum_1^k (x_i - \bar{x})^2 &= \sum_1^k (x_i - \bar{x})\, y_i. \end{aligned} \tag{2.4}$$

Scanned with OKEN Scanner

Deoarece $\sum_1^k (x_i - \bar{x}) = 0$, sistemul (2.4) are soluțiile

$$a = \frac{1}{k}\sum_1^k y_i = \bar{y}, \tag{2.5}$$

$$b = \frac{\sum_1^k (x_i - \bar{x})\, y_i}{\sum_1^k (x_i - \bar{x})^2}. \tag{2.6}$$

Ținînd seama că $\sum_1^k (x_i - \bar{x}) = 0$, urmează că

$$\sum_1^k (x_i - \bar{x})\, y_i = \sum_1^k (x_i - \bar{x})\, y_i - \bar{y}\sum_1^k (x_i - \bar{x}) =$$

$$= \sum_1^k (x_i - \bar{x})\,(y_i - \bar{y}),$$

ceea ce ne permite să scriem (2.6) sub forma

$$b = \frac{\sum_1^k (x_i - \bar{x})\,(y_i - \bar{y})}{\sum_1^k (x_i - \bar{x})^2}. \tag{2.6'}$$

Presupunem că $M(y_i) = \mu_i$ dată de (2.1) cu $x = x_i$. Avem

$$M(b) = \frac{\sum_1^k (x_i - \bar{x})\, M(y_i)}{\sum_1^k (x_i - \bar{x})^2} = \frac{\sum_1^k (x_i - \bar{x})\,[\alpha + \beta(x_i - \bar{x})]}{\sum_1^k (x_i - \bar{x})^2} =$$

$$= \alpha \frac{\sum_1^k (x_i - \bar{x})}{\sum_1^k (x_i - \bar{x})^2} + \beta \frac{\sum_1^k (x_i - \bar{x})^2}{\sum_1^k (x_i - \bar{x})^2} = \beta, \tag{2.7}$$

Scanned with OKEN Scanner

ceea ce ne permite să afirmăm că b este o estimație nedeplasată pentru β.

Deoarece am presupus că y_i sînt independente și repartizate normal cu dispersia σ^2, urmează că

$$D^2(a) = \frac{1}{k^2}\sum_1^k D^2(y_i) = \frac{1}{k^2}\, k\sigma^2 = \frac{\sigma^2}{k}, \tag{2.8}$$

$$D^2(b) = \frac{\sum_1^k (x_i - \bar{x})^2}{\left[\sum_1^k (x_i - \bar{x})^2\right]^2} D^2(y_i) = \frac{\sigma^2}{\sum_1^k (x_i - \bar{x})^2}. \tag{2.9}$$

Vom arăta că estimația de cel mai mic pătrat, dacă există, are cea mai mică dispersie. Presupunem că mai există o estimație lineară b' pentru β. Fie

$$b' = \sum_1^k c_i y_i. \tag{2.10}$$

Avem

$$M(b') = \sum_1^k c_i M(y_i) = \sum_1^k c_i[\alpha + \beta(x_i - \bar{x})] =$$

$$= \alpha \sum_1^k c_i + \beta \sum_1^k (x_i - \bar{x})\, c_i. \tag{2.11}$$

Pentru ca b' să fie o estimație nedeplasată pentru β trebuie ca $M(b') = \beta$ și din (2.11) obținem condițiile

$$\sum_1^k c_i = 0,$$

$$\sum_1^k (x_i - \bar{x})\, c_i = 1. \tag{2.12}$$

Scanned with OKEN Scanner

Avem

$$D^2(b') = \sum_1^k c_i^2 D^2(y_i) = \sigma^2 \sum_1^k c_i^2 =$$

$$= \sigma^2 \sum_1^k \left[c_i - \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} + \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} \right]^2$$

şi cum

$$2\sigma^2 \sum_1^k \left[c_i - \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} \right] \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} =$$

$$= 2\sigma^2 \frac{\sum_1^k c_i (x_i - \bar{x})}{\sum_1^k (x_i - \bar{x})^2} - 2\sigma^2 \frac{\sum_1^k (x_i - \bar{x})^2}{\left[\sum_1^k (x_i - \bar{x})^2 \right]^2} = 0 \tag{2.13}$$

obţinem

$$D^2(b') = \sigma^2 \sum_1^k \left[c_i - \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} \right]^2 + \sigma^2 \frac{\sum_1^k (x_i - \bar{x})^2}{\left[\sum_1^k (x_i - \bar{x})^2 \right]^2} =$$

$$= \sigma^2 \sum_1^k \left[c_i - \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2} \right]^2 + \frac{\sigma^2}{\sum_1^k (x_i - \bar{x})^2} \tag{2.14}$$

în care ultimul termen este o constantă.

Prin urmare pentru a face minimă dispersia estimaţiei b', putem acţiona numai asupra primului termen din (2.14). Luînd

$$c_i = \frac{x_i - \bar{x}}{\sum_1^k (x_i - \bar{x})^2}, \tag{2.15}$$

$$D^2(b') = \frac{\sigma^2}{\sum_1^k (x_i - \bar{x})^2}$$

Scanned with OKEN Scanner

și (2.10) devine

$$b' = \sum_1^k c_i y_i = \frac{\sum_1^k (x_i - \bar{x})\, i}{\sum_1^k (x_i - \bar{x})^2},$$

care este tocmai estimația de cel mai mic pătrat definită în (2.6). Prin urmare $b = b'$ și b este o estimație liniară nedeplasată de minimă dispersie. Întîmplător, din (2.14) observăm că dacă c_i este dată de (2.15), $D^2(b) = \sigma^2 \Big/ \sum_1^k (x_i - \bar{x})^2$, valoare găsită în (2.9).

§ 3. Analiza dispersiei și regresiei

Vom considera analiza regresiei din punct de vedere al analizei dispersiei.

În acest scop vom arăta că a și b sînt independente una de cealaltă și de s^2, estimația lui σ^2.

Abaterile observațiilor y_i de la valorile date de dreapta de regresie (2.1) se scriu

$$\begin{aligned} y_i - \mu_i &= (y_i - Y_i) + (Y_i - \mu_i) = \\ &= (y_i - Y_i) + \{a + b(x_i - \bar{x}) - [\alpha + \beta(x_i - \bar{x})]\} = \\ &= (y_i - Y_i) + (a - \alpha) + (b - \beta)(x_i - \bar{x}). \end{aligned} \tag{3.1}$$

Ridicînd la pătrat ambii membri ai relației (3.1) și sumînd după i obținem

$$\begin{aligned} \sum_1^k (y_i - \mu_i)^2 = \sum_1^k (y_i - Y_i)^2 + k(a - \alpha)^2 + \\ + (b - \beta)^2 \sum_1^k (x_i - \bar{x})^2. \end{aligned} \tag{3.2}$$

$\sum_1^k (y_i - \mu_i)^2$ este o sumă de k pătrate de variabile aleatoare independente repartizate $N(0, \sigma^2)$, ceea ce ne permite să afirmăm că $\sum_1^k (y_i - \mu_i)^2$ este repartizată ca $\sigma^2 \chi^2(k)$. (3.2) a fost partiționat

Scanned with OKEN Scanner

în trei componente. A doua și a treia componentă conțin o singură variabilă a și respectiv b, aceste componente avînd fiecare un singur grad de libertate. Prima componentă conține k variabile $y_i - Y_i$, însă ținînd seama de condițiile

$$\frac{\partial S}{\partial a} = -2\sum_1^k [y_i - a - b(x_i - \bar{x})] = 0,$$

$$\frac{\partial S}{\partial b} = -2\sum_1^k [y_i - a - b(x_i - \bar{x})](x_i - \bar{x}) = 0,$$

urmează că avem două restricții și astfel prima componentă are $k - 2$ grade de libertate. După teorema lui Cochran, $\sum_1^k (y_i - \mu_i)^2$, $k(a - \alpha)^2$ și $(b - \beta)^2 \sum_1^k (x_i - \bar{x})^2$ sînt independente și repartizate ca $\sigma^2\chi^2$ cu k, 1 și respectiv 1 grad de libertate. Urmează că $\sum_1^k (y_i - Y_i)^2$ este repartizată ca $\sigma^2\chi^2(k - 2)$. Avem situația din tabelul 7.1.

Tabelul 7.1

abaterile	grad de libertate	suma pătratelor	3/2
1	2	3	4
lui a de la α	1	$k(a - \alpha)^2$	s_a^2
lui b de la β	1	$(b - \beta)^2 \sum_1^k (x_i - \bar{x})^2$	s_b^2
valorilor observate de la dreapta estimatoare	k—2	$\sum_1^k (y_i - Y_i)^2$	s^2
valorilor observate de la dreapta adevărată	k	$\sum_1^k (y_i - \mu_i)^2$	

Scanned with OKEN Scanner

Urmează că

$$M(s^2) = M = \left[\frac{\sum_1^k (y_i - Y_i)^2}{k-2}\right] = \frac{\sigma^2}{k-2}(\chi^2(k-1)) = \sigma^2, \tag{3.3}$$

ceea ce ne arată că s^2 este o estimație pentru σ^2 și este independentă de a și b. De asemenea, urmează că $k(a-\alpha)^2$ este repartizată ca $\sigma^2\chi^2(1)$, astfel încît

$$M(k(a-\alpha)) = M[\sigma^2\chi^2(1)] = \sigma^2$$

și

$$D^2(a) = D^2[(a-\alpha)^2] = \sigma^2/k. \tag{3.4}$$

Analog, obținem

$$D^2(b) = \sigma^2 \Big/ \sum_1^k (x_i - \bar{x})^2. \tag{3.5}$$

Avem

$D^2(Y) = D^2[a + b(x-\bar{x})] =$
$= D^2(a) + (x-\bar{x})^2 D^2(b) + 2(x-\bar{x})\operatorname{cov}(a, b)$ și cum a și b sînt independente

$$D^2(Y) = \sigma^2\left[\frac{1}{k} + \frac{(x-\bar{x})^2}{\sum_1^k (x_i - \bar{x})^2}\right]. \tag{3.6}$$

Valoarea medie a lui Y este $\alpha + \beta(x-\bar{x}) = \eta$. Y este funcție lineară de a și b, care la rîndul lor sînt funcții de y_i, repartizate normal. Urmează că și Y este repartizată normal. Deci

$$\frac{Y - M(Y)}{D(Y)} \sim t(k-2). \tag{3.7}$$

În (3.7) la numitor nu avem $D(Y)$ dat de (3.6), ci

$$\sqrt{s^2\left[\frac{1}{k} + \frac{(x-\bar{x})^2}{\sum_1^k (x_i - \bar{x})^2}\right]}$$

cu s^2 definit în (3.3).

Scanned with OKEN Scanner

§ 4. Compararea a două drepte de regresie

Presupunem că avem două mulțimi de observații

$$(x_{11}, y_{11}), (x_{12}, y_{12}), \ldots, (x_{1n_1}, y_{1n_1})$$

și

$$(x_{21}, y_{21}), (x_{22}, y_{22}), \ldots, (x_{2n_2}, y_{2n_2}).$$

Ne propunem să vedem dacă pentru ajustarea acestor date este suficientă o singură dreaptă de regresie, sau sînt necesare drepte de regresie separate. Vom considera dreptele

$$Y_1 = a_1 + b_1(x - \bar{x}_1),$$

(4.1)

$$Y_2 = a_2 + b_2(x - \bar{x}_2).$$

Dacă dreptele sînt identice, $s_i^2 (i = 1, 2)$, vor estima dispersia comună σ^2 și raportul lor va fi repartizat ca F. Vom pune la numărător cea mai mare dintre valorile s_i^2. Se va folosi un test bilateral. Dacă ipoteza nulă este respinsă, atunci dreptele diferă. Dacă ipoteza nulă a unei dispersii reziduale comune este acceptată,

$$s^2 = \frac{(n_1 - 2)\, s_1^2 + (n_2 - 2)\, s_2^2}{n_1 - 2 + n_2 - 2} \tag{4.2}$$

este estimația comună pentru σ^2.

Dacă ipoteza nulă $\beta_1 = \beta_2$ este adevărată, atunci $(b_1 - b_2)$ este repartizată normal de medie zero și dispersia

$$D^2(b_1 - b_2) = D^2(b_1) + D^2(b_2) = \frac{\sigma^2}{\sum_1^{n_1} (x_{1i} - \bar{x}_1)^2} + \frac{\sigma^2}{\sum_1^{n_1} (x_{2i} - \bar{x}_2)^2} \tag{4.3}$$

unde σ^2 este estimată de s^2 definită de (4.2).

Urmează că în ipoteza nulă

$$\frac{b_1 - b_2}{s \sqrt{\dfrac{1}{\sum_1^{n_1} (x_{1i} - \bar{x}_1)^2} + \dfrac{1}{\sum_1^{n_2} (x_{2i} - \bar{x}_2)^2}}} \sim t(n_1 + n_2 - 4). \tag{4.4}$$

Dacă respingem ipoteza nulă, atunci dreptele au pante diferite și deci sînt diferite. Dacă ipoteza nulă este acceptată, estimăm panta comună și verificăm dacă coincid sau sînt paralele.

Scanned with OKEN Scanner

Vom estima panta comună. Avem două ecuații adevărate pentru două drepte paralele

$$\mu_1 = \alpha_1 + \beta(x - \bar{x}_1), \quad \mu_2 = \alpha_2 + \beta(x - \bar{x}_2) \tag{4.5}$$

și două ecuații estimatoare

$$Y_1 = a_1 + b(x - \bar{x}_1), \quad Y_2 = a_2 + b(x - \bar{x}_2). \tag{4.6}$$

Suma pătratelor abaterilor observațiilor de la cele două drepte estimatoare este

$$S = \sum_1^{n_1} (y_{1i} - Y_{1i})^2 + \sum_1^{n_2} (y_{2i} - Y_{2i})^2 =$$

$$= \sum_1^{n_1} [y_{1i} - a_1 - b(x_{1i} - \bar{x}_1)]^2 + \sum_1^{n_2} [y_{2i} - a_2 - b(x_{2i} - \bar{x}_2)]^2. \tag{4.7}$$

Estimațiile de cel mai mic pătrat sînt acelea care fac pe S minimă. Obținem

$$\frac{\partial S}{\partial a_1} = -2 \sum_1^{n_1} [y_{1i} - a_1 - b(x_{1i} - \bar{x}_1)] = 0,$$

$$\frac{\partial S}{\partial a_2} = -2 \sum_1^{n_2} [y_{2i} - a_2 - b(x_{i2} - \bar{x}_2)] = 0,$$

$$\frac{\partial S}{\partial b} = -2 \sum_1^{n_1} [y_{1i} - a_1 - b(x_{1i} - \bar{x}_1)] (x_{1i} - \bar{x}_1) -$$

$$- 2 \sum_1^{n_2} [y_{2i} - a_2 - b(x_{2i} - \bar{x}_1)] (x_{2i} - \bar{x}_2) = 0.$$

Din primele două ecuații, obținem

$$a_1 = \frac{1}{n_1} \sum_1^{n_1} y_{1i}, \qquad a_2 = \frac{1}{n_2} \sum_1^{n_2} y_{2i}, \tag{4.8}$$

iar dintr-a treia

$$b = \frac{\sum_1^{n_1} y_{1i} (x_{1i} - \bar{x}_1) + \sum_1^{n_2} y_{2i} (x_{2i} - \bar{x}_2)}{\sum_1^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_1^{n_2} (x_{2i} - \bar{x}_2)^2}. \tag{4.9}$$

Scanned with OKEN Scanner

Avem

$$D^2(b) = \frac{\sum_1^{n_1}(x_{1i}-\bar{x}_1)^2 D^2(y_{1i}) + \sum_1^{n_1}(x_{2i}-\bar{x}_2)^2 D^2(y_{2i})}{\left[\sum_{j=1}^{2}\sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2\right]^2} =$$

$$= \frac{\sigma^2}{\sum_{j=1}^{2}\sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2}. \tag{4.10}$$

Ținînd seama în (4.7) de (4.8) și (4.9) obținem

$$S = \left[\sum_{j=1}^{2}\sum_{i=1}^{n_j} y_{ji}^2 - \sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j} y_{ji}\right] - \frac{\left[\sum_{j=1}^{2}\sum_{i=1}^{n_j} y_{ji}(x_{ji}-\bar{x}_j)\right]^2}{\sum_{j=1}^{2}\sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2}. \tag{4.11}$$

Deoarece am estimat trei parametri cu ajutorul observațiilor, această sumă are $n_1 + n_2 - 3$ grade de libertate.

Verificăm ipoteza conform căreia cele două drepte paralele sînt identice. Dacă dreptele adevărate (4.5) sînt identice, atunci $\mu_1 = \mu_2$ pentru orice x și deci

$$\alpha_1 - \beta\bar{x}_1 = \alpha_2 - \beta\bar{x}_2$$

sau

$$(\alpha_1 - \alpha_2) - \beta(\bar{x}_1 - \bar{x}_2) = 0.$$

Prin urmare, $[(a_1 - a_2) - b(\bar{x}_1 - \bar{x}_2)]$ este repartizată normal, de medie

$$M[(a_1 - a_2) - b(\bar{x}_1 - \bar{x}_2)] = 0$$

și dispersie

$$D^2[(a_1 - a_2) - b(\bar{x}_1 - \bar{x}_2)] = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} + (\bar{x}_1 - \bar{x}_2)^2 D^2(b) =$$

$$= \sigma^2\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{\sum_{j=1}^{2}\sum_{i=1}^{n_j}(x_{ji}-\bar{x}_j)^2}\right].$$

Scanned with OKEN Scanner

Astfel, dacă dreptele sînt identice

$$\frac{(a_1 - a_2) - b(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{(\bar{x}_1 - \bar{x}_2)^2}{\sum_{j=1}^{2}\sum_{i=1}^{n_j}(\bar{x}_{ji} - \bar{x}_j)^2}}} \sim t(n_1 + n_2 - 3),$$

unde s se obține din (4.11).

Prin urmare compararea a două drepte de regresie comportă trei etape:

1) se verifică ipoteza conform căreia împrăștierea față de cele două drepte poate fi considerată ca omogenă;

2) se verifică ipoteza conform căreia dreptele paralele pot fi considerate ca o ajustare acceptabilă;

3) se verifică ipoteza conform căreia aceste două drepte paralele se confundă.

§ 5. Regresie lineară trecînd prin origine

Adeseori dreapta de regresie trece prin origine, adică dreapta de regresie adevărată este de forma

$$\mu = \beta x. \tag{5.1}$$

În cele ce urmează ne propunem să ajustăm o dreaptă de forma (5.1). Presupunem că y este repartizată normal de medie μ (dat de (5.1)) și dispersie σ^2 și că observațiile sînt independente. Ecuația dreptei de regresie estimatoare este (5.2)

$$Y = bx.$$

Suma pătratelor abaterilor dintre valorile observate y_i și valorile rezultate Y_i este

$$S = \Sigma(y_i - Y_i)^2 = \Sigma(y_i - bx_i)^2. \tag{5.3}$$

Valoarea lui b, care face minimă (5.3), este soluția ecuației

$$\frac{dS}{db} = -2\Sigma(y_i - bx_i)\,x_i = 0,$$

Scanned with OKEN Scanner

de unde

$$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2}. \tag{5.4}$$

Avem

$$D^2(b) = \frac{1}{(\Sigma x_i^2)^2} \Sigma x_i^2 D^2(y_i) = \frac{\sigma^2}{\Sigma x_i^2}, \tag{5.5}$$

$$D^2(Y) = x^2 D^2(b) = \frac{x^2\sigma^2}{\Sigma x_i^2}. \tag{5.6}$$

Putem scrie

$$y_i - \mu_i = (y_i - Y_i) + (Y_i - \mu_i) = (y_i - Y_i) + (b - \beta)\, x_i,$$

de unde ridicînd la pătrat și sumînd după i, obținem

$$\Sigma(y_i - \mu_i)^2 = \Sigma(y_i - Y_i)^2 + (b - \beta)^2 \Sigma x_i^2, \tag{5.7}$$

termenii fiind repartizați ca $\sigma^2\chi^2$ cu k, $k - 1$, respectiv un grad de libertate. Urmează că

$$s^2 = \frac{1}{k-1} \Sigma(y_i - Y_i)^2 \tag{5.8}$$

are media σ^2 și este independentă de b.

Putem scrie

$$y_i = (y_i - Y_i) + Y_i = (y_i - Y_i) + bx_i,$$

de unde ridicînd la pătrat și sumînd obținem

$$\Sigma y_i^2 = \Sigma(y_i - Y_i)^2 + \Sigma(bx_i)^2 + 2\Sigma(y_i - Y_i)\, bx_i =$$

$$= \Sigma(y_i - Y_i)^2 - \Sigma(bx_i)^2 - 2\Sigma(bx_i)^2 + 2\Sigma bx_i y_i =$$

$$= \Sigma(y_i - Y_i)^2 - \Sigma(bx_i)^2 + 2\Sigma bx_i y_i,$$

și ținînd seama de (5.4) obținem

$$\Sigma y_i^2 = \Sigma(y_i - Y_i)^2 + \frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}, \tag{5.9}$$

de unde

$$\Sigma(y_i - Y_i)^2 = \Sigma y_i^2 - \frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}. \tag{5.10}$$

Scanned with OKEN Scanner

Se observă că

$$\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2} = \left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right)^2 \Sigma x_i^2 = b^2 \Sigma x_i^2$$

și din (5.7)

$$(b - \beta)^2 \Sigma x_i^2 = b^2 \Sigma x_i^2 + \beta^2 \Sigma x_i^2 - 2b\beta \Sigma x_i^2,$$

are media σ^2, astfel încît aplicînd media obținem

$$\sigma^2 = M\left[\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}\right] + \beta^2 \Sigma x_i^2 - 2\beta \Sigma x_i^2 M(b) =$$

$$= M\left[\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}\right] - \beta^2 \Sigma x_i^2.$$

Astfel

$$M\left[\frac{(\Sigma x_i y_i)^2}{\Sigma x_i^2}\right] = \sigma^2 + \beta^2 \Sigma x_i^2. \tag{5.11}$$

Putem construi o tabelă pentru analiza dispersiei corespunzătoare lui (5.9) (tabela 7.2).

Tabelul 7.2

abaterile	suma pătratelor	gr. de libertate	media
de la dreaptă	$(\Sigma x_i y_i)^2 / \Sigma x_i^2$	1	$\sigma^2 + \beta^2 \Sigma x_i^2$
reziduale	$\Sigma (y_i - Y_i)^2$	k—1	σ^2
totale față de origine	Σy_i^2	k	

§ 6. Regresie lineară cu mai multe observații y la fiecare x

În acest paragraf presupunem că avem observațiile

$$\begin{array}{ll} y_{11}, y_{12}, \ldots, y_{1n_1} & \text{pentru } x_1, \\ \vdots & \vdots \\ y_{i1}, y_{i2}, \ldots, y_{in_i} & \text{pentru } x_i, \\ \vdots & \vdots \\ y_{k1}, y_{k2}, \ldots, y_{kn_k} & \text{pentru } x_k. \end{array}$$

Scanned with OKEN Scanner

Fie

$$\bar{x} = \frac{\sum_1^k n_i x_i}{\sum_1^k n_i}, \qquad \bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \qquad s_{1i}^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2.$$

Se observă că $\bar{x}$ este media valorilor x_i ponderată cu numărul de observații în fiecare punct, iar s_{1i}^2 este dispersia în interiorul fiecărui grup de observații rezultate. Cu testul lui Bartlett, putem verifica ipoteza conform căreia s_{1i}^2 provin din populații cu aceeași dispersie σ^2. În practică, acesta nu aduce satisfacții deoarece n_i sînt adeseori mici și testul va avea o putere mică. De asemenea s_{1i}^2 sînt ordonate după mărimea x-lor pe baza cărora s-au calculat și testul lui Bartlett nu ține seama de această ordine.

Vom considera ca estimație a lui σ^2, cantitatea

$$(6.1) \qquad s_1^2 = \frac{\sum_1^k (n_i - 1)\, s^2}{\sum_1^k (n_i - 1)} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{\sum_1^k n_i - k}$$

Suma pătratelor abaterilor valorilor observate y_{ij} de la valorile estimate Y_i este

$$(6.2) \qquad S = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - Y_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij} - a - b(x_i - \bar{x})]^2.$$

Egalînd cu zero derivatele parțiale ale lui (6.2) în raport cu a și b, obținem

$$(6.3) \qquad \frac{\partial S}{\partial a} = -2 \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij} - a - b(x_i - \bar{x})] = 0,$$

$$(6.4) \qquad \frac{\partial S}{\partial b} = -2 \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij} - a - b(x_i - \bar{x})]\,(x_i - \bar{x}) = 0,$$

sau

$$(6.5) \qquad a \sum_1^k n_i + b \sum_1^k n_i (x_i - \bar{x}) = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij},$$

$$(6.6) \qquad a \sum_1^k n_i (x_i - \bar{x}) + b \sum_1^k n_i (x_i - \bar{x})^2 = \sum_1^k n_i (x_i - \bar{x})\, \bar{y}_i,$$

Scanned with OKEN Scanner

deoarece

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_i-\bar{x})\,y_{ij}=\sum_{i=1}^{k}(x_i-\bar{x})\sum_{j=1}^{n_i}y_{ij}=\sum_{i=1}^{k}n_i(x_i-\bar{x})\,\bar{y}_i.$$

Deoarece

$$\sum_{1}^{k}n_i(x_i-\bar{x})=\sum_{1}^{k}n_ix_i-\bar{x}\sum_{1}^{k}n_i=0,$$

din (6.5) obținem

$$a=\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}y_{ij}}{\sum_{1}^{k}n_i}=\frac{\sum_{i=1}^{k}n_i\bar{y}_i}{\sum_{1}^{k}n_i}=\bar{y}, \tag{6.7}$$

iar din (6.6)

$$b=\frac{\sum_{1}^{k}n_i(x_i-\bar{x})\,\bar{y}_i}{\sum_{1}^{k}n_i(x_i-\bar{x})^2}=\frac{\sum_{1}^{k}n_i(x_i-\bar{x})\,\bar{y}_i-\sum_{1}^{k}n_i(x_i-\bar{x})\,\bar{y}}{\sum_{1}^{k}n_i(x_i-\bar{x})^2}=$$

$$=\frac{\sum_{1}^{k}n_i(x_i-\bar{x})\,(\bar{y}_i-\bar{y})}{\sum_{1}^{k}n_i(x_i-\bar{x})^2}.$$

Vom considera acum o prezentare a celor anterioare în limbajul analizei dispersiei.

Abaterea observației y_{ij} de la valoarea μ_i dată de ecuația de regresie adevărată (2.1) poate fi scrisă

$$y_{ij}-\mu_i=(y_{ij}-\bar{y}_i)+(\bar{y}_i-\mu_i)=(y_{ij}-\bar{y}_i)+(\bar{y}_i-Y_i)+$$

$$+(Y_i-\mu_i)=(y_{ij}-\bar{y}_i)+(\bar{y}_i-Y_i)+(a-\alpha)+$$

$$+(b-\beta)\,(x_i-\bar{x}).$$

Scanned with OKEN Scanner

Semnificaţiile termenilor din membrul drept al ecuaţiei (6.9) sînt:

$(y_{ij} - \bar{y}_i)$ — abaterea unei observaţii y_{ij} de la medie $\bar{y}_i$ a n_i observaţii corespunzătoare la x_i;

$(\bar{y}_i - Y_i)$ — abaterea mediei $\bar{y}_i$ a n_i observaţii corespunzătoare la x_i de la valoarea Y_i dată de ecuţia estimatoare;

$(a - \alpha)$ — abaterea estimaţiei a de la valoarea adevărată α;

$(b - \beta)\ (x_i - \bar{x})$ — distanţa în direcţia lui y dată de abaterea estimaţiei b de la valoarea adevărată β.

Ridicînd la pătrat ambii membri ai relaţiei (6.9) şi apoi sumînd după i şi j obţinem

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{1}^{k} n_i(\bar{y}_i - Y_i)^2 +$$

$$+ (a - \alpha)^2 \sum_{1}^{k} n_i + (b - \beta)^2 \sum_{1}^{k} n_i(x_i - \bar{x})^2. \tag{6.10}$$

În (6.10) am desfăcut o sumă de pătrate cu $\sum_{1}^{k} n_i$ grade de libertate în patru sume de pătrate. A treia şi a patra sumă de pătrate conţin fiecare cîte o singură variabilă, a şi respectiv b, şi acestea au fiecare un grad de libertate. A doua componentă conţine k variabile, cărora, însă, li se impun condiţiile (6.3) şi (6.4), şi astfel are $(k - 2)$ grade de libertate. Prima componentă este numărătorul lui (6.1) care are $\left(\sum_{1}^{k} n_i - k\right)$ grade de libertate. Astfel suma sumelor de pătrate este egală cu suma pătratelor din membrul drept şi de asemenea suma gradelor de libertate. Din teorema lui Cochran urmează că sumele de pătrate sînt independente şi repartizate ca $\sigma^2\chi^2$. Ecuaţia (6.10) este trecută în tabelul 7.3.

Dacă $(a - \alpha)^2 \sum_{1}^{k} n_i$ este repartizată ca $\sigma^2\chi^2(1)$, atunci

$$D^2(a) = D^2[(a - \alpha)^2] = \frac{\sigma^2}{\sum_{1}^{k} n_i}, \tag{6.11}$$

$$D^2(b) = M[(b - M(b))^2] = \frac{\sigma^2}{\sum_{1}^{k} n_i(x_i - \bar{x})^2} \tag{6.12}$$

Tabelul 7.3

abaterea	suma pătratelor	nr. gradelor de libertate	2/3
1	2	3	4
în interiorul grupelor	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$	$\sum_{1}^{k} n_i - k$	s_1^2
între grupele de medii	$\sum_{1}^{k} n_i(\bar{y}_i - Y_i)^2$	$k-2$	s_2^2
de la a la α	$(a-\alpha)^2 \sum_{1}^{k} n_i$	1	s_3^2
de la b la β	$(b-\beta)^2 \sum_{1}^{k} n_i(x_i-\bar{x})^2$	1	s_4^2
față de linia teoretică	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2$	$\sum_{1}^{k} n_i$	

Deoarece a și b sînt independente

$$D^2(Y) = D^2[a + b(x - \bar{x})] = D^2(a) + (x - \bar{x})^2\, D^2(b) = \tag{6.13}$$

$$= \sigma^2\left[\frac{1}{\sum_{1}^{k} n_i} + \frac{(x-\bar{x})^2}{\sum_{1}^{k} n_i(x_i-\bar{x})^2}\right].$$

Putem scrie

$$y_{ij} - \bar{y} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - Y_i) + (Y_i - \bar{y})$$

de unde,

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 +$$

$$+\sum_{1}^{k} n_i(\bar{y}_i - Y_i)^2 + \sum_{1}^{k} n_i(Y_i - \bar{y})^2. \tag{6.14}$$

Primul termen din membrul drept al lui (6.14) este numărătorul lui s_1^2 definit în (6.1), despre care știm că este repartizat ca $\sigma^2\chi^2$ cu $\left(\sum_1^k n_i - k\right)$ grade de libertate; $M(s_1^2) = \sigma^2$.

Al doilea termen este numărătorul lui s_2^2 din tabelul 7.3, care în presupunerea că μ_i se găsește pe dreapta

$$\alpha + \beta(x_i - \bar{x}),$$

este repartizat ca $\sigma^2\chi^2$, cu $(k-2)$ grade de libertate. Este posibil ca μ_i să nu varieze pe această dreaptă. Chiar dacă μ_i nu variază pe o linie dreaptă, prin metoda celor mai mici pătrate, este posibil să se ajusteze o dreaptă la punctele (x_i, μ_i) și analog cu (6.7) și (6.8), avem parametrii

$$\alpha = \bar{\mu} = \frac{\sum_1^k n_i\mu_i}{\sum_1^k n_i}, \qquad \beta = \frac{\sum_1^k n_i(x_i - \bar{x})\,\mu_i}{\sum_1^k n_i(x_i - \bar{x})^2}.$$

Putem scrie

$$\bar{y}_i - \mu_i = (\bar{y}_i - Y_i) + (Y_i - \mu_i) =$$

$$= (\bar{y}_i - Y_i) + [a + b(x_i - \bar{x}) - \bar{\mu} - \beta(x_i - \bar{x}_i)] - [\mu_i -$$

$$- \bar{\mu} - \beta(x_i - \bar{x})] = (\bar{y}_i - Y_i) - [\mu_i - \bar{\mu} -$$

$$-\beta(x_i - \bar{x})] + (a - \bar{\mu}) + (b - \beta)\,(x_i - \bar{x}).$$

Ridicînd la pătrat și sumînd după i și j obținem

$$\sum_1^k n_i(\bar{y}_i - \mu_i)^2 = \sum_1^k n_i\{(\bar{y}_i - Y_i) - [\mu_i - \mu -$$

(6.15)

$$- \beta(x_i - \bar{x})]\}^2 + (a - \bar{\mu})^2 \sum_1^k n_i + (b - \beta)^2 \sum_1^k n_i(x_i - \bar{x})^2.$$

Membrul stîng este repartizat ca $\sigma^2\chi^2$ cu k grade de libertate iar termenii din dreapta vor fi repartizați ca $\sigma^2\chi^2$ cu $(k-2)$, 1, respectiv 1 grad de libertate.

Scanned with OKEN Scanner

Avem

$$M\left(\sum_1^k n_i\{(\bar{y}_i - Y_i) - [\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]\}^2\right) =$$

$$= (k-2)\,\sigma^2 = M\left(\sum_1^k n_i(\bar{y}_i - Y_i)^2\right) + M\left(\sum_1^k n_i[\mu_i -\right.$$

$$\left. -\bar{\mu} - \beta(x_i - \bar{x})]^2\right) - 2M\left(\sum_1^k n_i(\bar{y}_i - Y_i)\,[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]\right)$$

și cum

$$M\left(\sum_1^k n_i(\bar{y}_i - Y_i)\,[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]\right) =$$

$$= \sum_1^k n_i[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]\,M(\bar{y}_i - Y_i) =$$

$$= \sum_1^k n_i[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})][\mu_i - \alpha - \beta(x_i - \bar{x})] =$$

$$= \sum_1^k n_i[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]^2,$$

deoarece $\alpha = \bar{\eta}$, din (6.16) obținem

$$M\left(\sum_1^k n_i(y_i - Y_i)^2\right) = (k-2)\sigma^2 +$$

(6.17)

$$\sum_1^k n_i[\mu_i - \bar{\mu} - \beta(x_i - \bar{x})]^2.$$

De asemenea

$$M\left(\sum_1^k n_i(Y_i - \bar{y})^2\right) = \sigma^2 + \beta^2 \sum_1^k n_i(x_i - \bar{x})^2. \tag{6.18}$$

Ținînd seama de rezultatele obținute, (6.14) se pune sub forma din tabelul 7.4.

Scanned with OKEN Scanner

Tabelul 7.4

abaterea	suma pătratelor	nr. gradelor de libertate	2/3	$M(2/3)$
1	2	3	4	5
în interiorul grupurilor	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$	$\sum_{1}^{k} n_i - k$	s_1^2	σ^2
mediei adevărate de grup față de dreaptă	$\sum_{1}^{k} n_i(\bar{y}_i - Y_i)^2$	$k-2$	s_2^2	$\sigma^2 + \frac{\sum_{1}^{k} n_i[\mu_i - \bar{\mu} - \beta(x_i-\bar{x})]^2}{k-2}$
panta dreptei	$\sum_{1}^{k} n_i(Y_i - \bar{y})^2$	1	s_3^2	$\sigma^2 + \beta^2\sum_{1}^{k} n_i(x_i-\bar{x})^2$
Total	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$	$\sum_{1}^{k} n_i - 1$		

Pentru a verifica ipoteza conform căreia μ_i variază pe o linie dreaptă, vom folosi raportul s_2^2/s_3^2, deoarece dacă este adevărată ipoteza nulă, $M(s_2^2) = \sigma^2$ și s_2^2 este independentă de s_3^2.

Pentru același motiv, în cazul în care dorim să verificăm ipoteza conform căreia $\beta = 0$, folosim raportul s_1^2/s_3^2.

Scanned with OKEN Scanner

CAPITOLUL

8

REPARTIȚIA NORMALĂ BIDIMENSIONALĂ ȘI COEFICIENTUL DE CORELAȚIE

§ 1. Repartiția normală bidimensională

Această repartiție este caracterizată de densitatea

$$f(x, y) =$$

(1.1)

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \right.\right.$$

$$\left.\left. - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\},$$

$$-\infty < x, y < \infty, \quad -\infty < \mu_x, \mu_y < \infty,$$

$$\sigma_x, \sigma_y > 0, \quad -1 < \rho < 1.$$

În cazul în care

$$\mu_x = \mu_y = 0, \quad \sigma_x = \sigma_y = 1,$$

(1.1) capătă forma

$$(1.2)\quad f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\}$$

și caracterizează așa-numita *repartiție normală bidimensională standard* (sau *redusă*) cu coeficientul de corelație ρ.

$f(x, y)$, dată de (1.1), este $\geqslant 0$ și continuă în amîndouă variabilele.

Repartiția marginală a variabilei aleatoare Y, în repartiția avînd densitatea (1.1), este caracterizată de densitatea $f(y)$. Punînd

$$u = \frac{x-\mu_x}{\sigma_x}, \; v = \frac{y-\mu_y}{\sigma_y}, \quad w = \frac{u-\rho u}{\sqrt{1-\rho^2}},$$

avem succesiv

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\times \int \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right\}\sigma_x\, du =$$

$$= \frac{1}{2\pi\sigma_y}\int_{-\infty}^{\infty} \exp\left(-\frac{w^2+v^2}{2}\right) dw =$$

$$= \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{v^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{w^2}{2}}\, dw = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{v^2}{2}}$$

sau

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]. \tag{1.3}$$

Analog, obținem

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]. \tag{1.4}$$

Deoarece densitățile marginale (1.3) și (1.4) sînt densități corespunzătoare unor variabile normale unidimensionale, obținem

$$M(X) = \mu_x, \quad M(Y) = \mu_y, \quad D^2(X) = \sigma_x^2, \quad D^2(Y) = \sigma_y^2.$$

Pentru covarianța variabilelor X și Y avem succesiv

$$\operatorname{cov}(X, Y) = \iint (x - \mu_x)(y - \mu_y) f(x, y)\, dx\, dy =$$

$$= \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}} \iint uv \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right\} du\, dv =$$

$$= \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}} \iint uv \exp\left\{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2 + v^2(1-\rho^2)\right]\right\} du\, dv =$$

$$= \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}} \int_{-\infty}^{\infty} u \exp\left\{-\frac{1}{2(1-\rho^2)}\left[(u-\rho v)^2\right]\right\} du\, dv,$$

Scanned with OKEN Scanner

sau

$$\operatorname{cov}(X, Y) = \frac{\sigma_x \sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}} \int_{-\infty}^{\infty} (u - \rho v) \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2]\right\} du\, dv +$$

(1.5)

$$+ \frac{\sigma_x \sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} v e^{-\frac{v^2}{2}} \int_{-\infty}^{\infty} \rho v \exp\left\{-\frac{1}{2(1-\rho^2)}[(u-\rho v)^2]\right\} du\, dv.$$

Punînd $w_x = u - \rho v$, în integrala interioară din primul termen al relației (1.5), se obține o integrală similară cu aceea folosită pentru determinarea mediei unei variabile aleatoare repartizate $N(0, 1-\rho^2)$.

Prin urmare primul termen este zero. Astfel

$$\operatorname{cov}(X, Y) =$$

$$= \frac{\sigma_x \sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \rho v^2 e^{-\frac{v^2}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2\right\} du\, dv =$$

$$= \frac{\sigma_x \sigma_y}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \rho v^2 e^{-\frac{v^2}{2}} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)} w_x^2\right\} dw\, dv =$$

$$= \frac{\sigma_x \sigma_y}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \rho v^2 e^{-\frac{v^2}{2}} dv = \rho \sigma_x \sigma_y.$$

Deoarece $\sigma_x, \sigma_y > 0$, covarianța este zero dacă și numai dacă $\rho = 0$. În acest caz (1.1) devine

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} =$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right] \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right] =$$

$$= f(x) f(y),$$

ceea ce ne arată că X și Y sînt variabile aleatoare independente. În acest caz particular (repartiția normală bidimensională) X și Y sînt independente dacă și numai dacă $\operatorname{cov}(X, Y) = 0$, lucru care în general nu este adevărat.

Scanned with OKEN Scanner

Natura suprafeței generate de repartiția normală bidimensională se poate vedea din (1.1). Dacă egalăm (1.1) cu o constantă aceasta implică

$$\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = \text{const.}$$

Această ecuație, pentru diferite valori ale constantei, definește elipse. Cînd $\rho = 0$, această ecuație devine

$$\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = \text{const.},$$

care definesc elipse ale căror axe principale sînt paralele cu axele Ox și Oy.

Ținînd seama de (1.1), (1.3) și de definiția densității condiționate, obținem

$$f(x \mid y) = = \frac{f(x, y)}{f(y)} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \right.\right.$$

$$\left.\left. + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \right.\right.$$

$$\left.\left. - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \rho^2\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} =$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \rho\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right.,$$

sau

$$f(x \mid y) = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left(\left[x - \right.\right.\right.$$

(1.6)

$$\left.\left.\left. - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)\right]^2\right\},$$

Scanned with OKEN Scanner

ceea ce ne arată că avem de-a face cu o repartiţie normală de medie

$$M(X \mid y) = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y) \tag{1.7}$$

şi dispersie

$$D^2(X \mid y) = \sigma_x^2(1 - \rho^2). \tag{1.8}$$

Din simetria densităţii, obţinem

$$f(y \mid x) = \frac{1}{\sigma_y \sqrt{2\pi} \sqrt{1 - \rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)\sigma_y^2} \left[y - (\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)) \right]^2 \right\}, \tag{1.9}$$

cu

$$M(Y \mid x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x) \tag{1.10}$$

şi

$$D^2(Y \mid x) = \sigma_y^2(1 - \rho^2). \tag{1.11}$$

Mediile repartiţiilor condiţionate sînt funcţii lineare de cealaltă variabilă, iar dispersiile sînt constante. Reamintim că în capitolul 7, § 2 modelul pentru regresia liniară presupune y repartizată normal cu media funcţie lineară de cealaltă variabilă x şi dispersia constantă, deci putem folosi rezultatele din capitolul 7.

În cazul de faţă avem două drepte de regresie (1.7) şi (1.8), cu coeficienţii de regresie

$$\beta_{x|y} = \rho \frac{\sigma_x}{\sigma_y}, \quad \beta_{Y|x} = \rho \frac{\sigma_y}{\sigma_x} \tag{1.12}$$

Produsul acestor coeficienţi este ρ^2.

Punctul de intersecţie al dreptelor de regresie este punctul (μ_x, μ_y).

Dacă notăm prin $\hat{B}$ unghiul dintre aceste două drepte de regresie avem

$$\operatorname{tg} \hat{B} = \frac{1 - \rho^2}{\rho} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \tag{1.13}$$

Scanned with OKEN Scanner

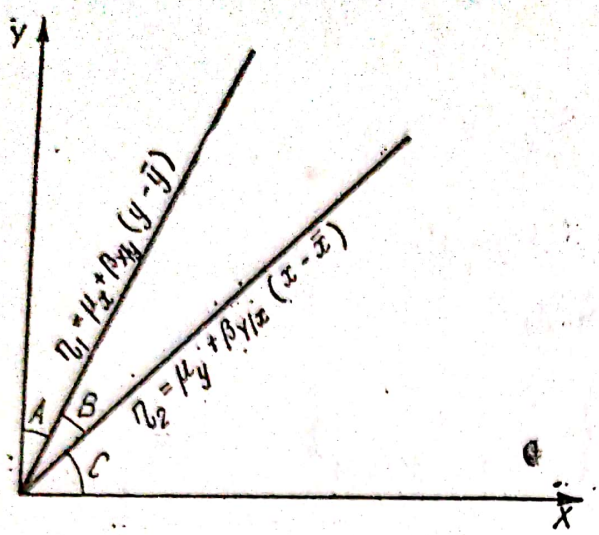


Fig. 8.1

Într-adevăr, dacă considerăm situaţia din figura 8.1,

$$\beta_{X|y} = \operatorname{tg} \hat{A}, \qquad \beta_{Y|x} = \operatorname{tg} \hat{C}$$

şi ţinînd seama că

$$\hat{A} + (\hat{B} + 90°) + \hat{C} = 180°,$$

implică

$$\operatorname{tg} \hat{A} + \operatorname{tg} (\hat{B} + 90°) + \operatorname{tg} \hat{C} = \operatorname{tg} \hat{A} (\operatorname{tg} \hat{B} + 90°) \operatorname{tg} \hat{C},$$

obţinem

$$\operatorname{tg} (\hat{B} + 90°) = \frac{\operatorname{tg} \hat{A} + \operatorname{tg} \hat{C}}{\operatorname{tg} \hat{A} \operatorname{tg} \hat{C} - 1} = \frac{\rho \frac{\sigma_x}{\sigma_y} + \rho \frac{\sigma_y}{\sigma_x}}{\rho^2 - 1} = \frac{\rho}{\rho^2 - 1} \cdot \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x \sigma_y},$$

astfel încît

$$\operatorname{tg} \hat{B} = - \frac{1}{\operatorname{tg} (\hat{B} + 90°)} = \frac{1 - \rho^2}{\rho} \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2},$$

ceea ce confirmă afirmaţia conţinută în (1.13).

Cînd $\rho = 0$, $\operatorname{tg} \hat{B} = \infty$ şi dreptele de regresie vor fi perpendiculare.

Cînd $\rho = 1$, $\operatorname{tg} \hat{B} = 0$ şi dreptele de regresie vor fi identice.

Din (1.8) şi (1.11), obţinem

$$\frac{D^2(X \mid y)}{D^2(X)} = 1 - \rho^2 = \frac{D^2(Y \mid x)}{D^2(Y)}, \tag{1.14}$$

ceea ce ne arată că

$$\rho = 0 \Rightarrow D^2(X \mid y) = D^2(X), \quad D^2(Y \mid x) = D^2(Y),$$

$$\rho = 1 \Rightarrow D^2(X \mid y) = 0.$$

Astfel ρ^2 poate fi privit ca măsură a fracţiei împrăştierii lui Y datorită regresiei asupra lui X şi invers.

Scanned with OKEN Scanner

§ 2. Estimarea parametrilor repartiției normale bidimensionale

Fie

$$(2.1)\qquad s_x^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2, \qquad s_y^2 = \frac{1}{n}\sum_1^n (y_j - \bar{y})^2,$$

$$(2.2)\qquad s_{xy} = \frac{1}{n}\sum_1^n (x_i - \bar{x})(y_i - \bar{y}),$$

(2.3) $P = \prod_1^n f(x_i, y_i)$ — funcția de verosimilitate corespunzătoare observațiilor (x_i, y_i) $1 \leqslant i \leqslant n$, dintr-o populație caracterizată de un vector aleator (X, Y) avînd densitatea dată de (1.1).

Avem

$$P = \left(\frac{1}{2\sigma_x\sigma_y\sqrt{1-\rho^2}}\right)^n \exp\left\{-\frac{1}{2(1-\rho^2)}\sum\left[\left(\frac{x_i-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x_i-\mu_x}{\sigma_x}\right)\left(\frac{y_i-\mu_y}{\sigma_y}\right) + \left(\frac{y_i-\mu_y}{\sigma_y}\right)^2\right]\right\},$$

de unde *

$$\frac{\partial \ln P}{\partial \mu_x} = 0 \Rightarrow -\frac{1}{\sigma_x^2}\sum_1^n (x_i - \hat{\mu}_x) + \frac{\rho}{\sigma_x\sigma_y}\sum_1^n (y_i - \hat{\mu}_y) = 0,$$

$$\frac{\partial \ln P}{\partial \mu_y} = 0 \Rightarrow \frac{\rho}{\sigma_x\sigma_y}\sum_1^n (x_i - \hat{\mu}_x) - \frac{1}{\sigma_y^2}\sum_1^n (y_i - \hat{\mu}_y) = 0$$

sistem ale cărui soluții sînt

$$(2.4)\qquad \hat{\mu}_x = \frac{1}{n}\sum_1^n x_i = \bar{x}, \qquad \hat{\mu}_y = \frac{1}{n}\sum_1^n y_i = \bar{y}$$

$$(2.5)\qquad \frac{\partial \ln P}{\partial \sigma_z} = 0 \Rightarrow \quad \frac{1}{\hat{\sigma}_x^2}\sum_1^n (x_i - \hat{\mu}_x)^2 = \frac{1}{\hat{\sigma}_x\hat{\sigma}_y}\sum_1^n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) + n(1 - \hat{\rho}^2).$$

* $\hat{}$ reprezintă estimația de maximă verosimilitate a parametrului.

Scanned with OKEN Scanner

Dacă în (2.5) ţinem seama de (2.1), (2.2) şi (2.4), obţinem

$$\frac{s_x^2}{\hat{\sigma}_x^2} = \hat{\rho}\,\frac{s_{xy}}{\hat{\sigma}_x\hat{\sigma}_y} + (1 - \hat{\rho}^2). \tag{2.6}$$

Din

$$\frac{\partial \ln P}{\partial \sigma_y} = 0,$$

obţinem

$$\frac{s_y^2}{\hat{\sigma}_y^2} = \hat{\rho}\,\frac{s_{xy}}{\hat{\sigma}_x\hat{\sigma}_y} + (1 - \hat{\rho}^2). \tag{2.7}$$

Din (2.6) şi (2.7) obţinem

$$\hat{\sigma}_2 = \hat{\sigma}_1\frac{s_y}{s_x}. \tag{2.8}$$

Ţinînd seama în (2.6) de (2.8) şi rezolvînd-o în raport cu $\hat{\sigma}_1^2$, obţinem

$$\hat{\sigma}_x^2 = \frac{s_x}{s_y}\cdot\frac{s_x s_y - \hat{\rho}\,s_{xy}}{1 - \hat{\rho}^2} \tag{2.9}$$

şi analog

$$\hat{\sigma}_y^2 = \frac{s_y}{s_x}\cdot\frac{s_x s_y - \hat{\rho}\,s_{xy}}{1 - \hat{\rho}^2}. \tag{2.10}$$

(2.9) şi (2.10) conţin pe $\hat{\rho}^2$ care este însă necunoscut. Îl obţinem pe ρ^2 ca soluţie a ecuaţiei

$$\frac{\partial \ln P}{\partial \rho} = 0.$$

Avem

$$\frac{\partial \ln P}{\partial \rho} = \frac{n\rho}{1-\rho^2} + \frac{1}{1-\rho^2}\sum_1^n \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x\sigma_y} - \tag{2.11}$$

$$- \frac{\rho}{(1-\rho^2)^2}\left[\sum_1^n \frac{(x_i - \mu_x)^2}{\sigma_x^2} - 2\rho\sum_1^n \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x\sigma_y} + \sum_1^n \frac{(y_i - \mu_y)^2}{\sigma_y^2}\right].$$

Scanned with OKEN Scanner

Egalînd (2.11) cu zero, înlocuind μ_x și μ_y prin $\bar{x}$ și $\bar{y}$, ținînd seama de (2.1) și (2.2) și multiplicînd prin $(1 - \hat{\rho}^2)/n$, obținem

$$\hat{\rho} + \frac{s_{12}}{\hat{\sigma}_x \hat{\sigma}_y} - \frac{\hat{\rho}}{1 - \hat{\rho}^2}\left(\frac{s_y^2}{\hat{\sigma}_x^2} - 2\hat{\rho}\,\frac{s_x s_y}{\hat{\sigma}_x \hat{\sigma}_y} + \frac{s_y^2}{\hat{\sigma}_y^2}\right) = 0. \tag{2.12}$$

Înlocuind în (3.12) $\hat{\sigma}_x$ și $\hat{\sigma}_y$ prin valorile date în (2.9) și (2.10) găsim estimația $\hat{\rho}^2$ de maximă verosimilitate pentru ρ.

$$\hat{\rho} = \frac{s_{12}}{\sqrt{s_1^2 s_2^2}}. \tag{2.13}$$

Ținînd seama de această valoare în (2.9) și (2.10) obținem estimațiile de maximă verosimilitate pentru σ_x^2 și σ_y^2.

$$\hat{\sigma}_x^2 = s_x^2, \quad \hat{\sigma}_y^2 = s_y^2. \tag{2.14}$$

Putem scrie

$$ns_{xy} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) =$$

$$= \sum_1^n (x_i - \mu_x)(y_i - \mu_y) - n(\bar{x} - \mu_x)(\bar{y} - \mu_y),$$

de unde

$$M[ns_{xy}] = n \operatorname{cov}(X, Y) - n \operatorname{cov}(\bar{x}, \bar{y})$$

și cum

$$\operatorname{cov}(\bar{x}, \bar{y}) = \frac{1}{n} \operatorname{cov}(X, Y),$$

$$M(s_{xy}) = \frac{n-1}{n} \operatorname{cov}(X, Y)$$

ceea ce arată că s_{xy} definit în (2.2) este o estimație deplasată pentru cov (X, Y).

O estimație nedeplasată pentru cov (X, Y) este dată de

$$s'_{xy} = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})(y_i - \bar{y}). \tag{2.2'}$$

Scanned with OKEN Scanner

Estimația de maximă verosimilitate pentru ρ, se notează prin r, are expresia

$$r = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2}} \tag{2.15}$$

și poartă numele de coeficient de corelație de selecție.

Se poate arăta că

$$M(r) = \rho - \frac{\rho(1-\rho^2)}{2n} + \ldots,$$

ceea ce ne arată că r este o estimație deplasată pentru ρ.

Coeficientul de regresie $\beta_{Y|x}$ se estimează de $b_{Y|x} = r\frac{s_y}{s_x}$. Dacă în expresia lui $b_{Y|x}$ ținem seama de (2.15) și de expresiile lui s_y și s_x, obținem

$$b_{Y|x} = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2} \tag{2.16}$$

care este identică cu 7.2.6′.

Estimația pentru dispersia reziduală față de linia de regresie este

$$s^2_{Y|x} = \frac{1}{n-2}\left\{\sum_1^n (y_i - \bar{y})^2 - \frac{\left[\sum_1^n (y_i - \bar{y})(x_i - \bar{x})\right]^2}{\sum_1^n (x_i - \bar{x})^2}\right\} =$$

$$= \frac{n-1}{n-2}\left[\frac{\sum_1^n (y_i - \bar{y})^2}{n-1}\right]\left\{1 - \left[\frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2}}\right]^2\right\} =$$

$$= \frac{n-1}{n-2} s_y^2(1 - r^2).$$

Scanned with OKEN Scanner

§ 3. Teste de semnificaţie pentru coeficientul de corelaţie

Ştim că b este repartizat normal de medie β şi dispersia dată de 7.2.9. Urmează că

$$\frac{b_{Y|x} - \beta_{Y|x}}{s_{Y|x} \Big/ \sqrt{\sum_1^n (x_i - \bar{x})^2}} \sim t(n-2). \tag{3.1}$$

Ipoteza conform căreia $\beta_{Y|x} = 0$ se verifică folosind statistica

$$\frac{b_{Y|x}}{s_{Y|x} \Big/ \sqrt{\sum_1^n (x_i - \bar{x})^2}} = \frac{r \dfrac{s_y}{s_x}}{\sqrt{\dfrac{n-1}{n-2} s_y^2 (1 - r^2)/(n-1)s_x^2}} = \tag{3.2}$$

$$= \sqrt{n-2}\, \frac{r}{\sqrt{1 - r^2}},$$

care în ipoteza nulă este repartizată Student cu $n - 2$ grade de libertate.

Însă ipoteza $\beta_{Y|x} = 0$ este aceeaşi cu ipoteza $\rho = 0$.

Fisher arată că

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r} = 1{,}1513 \lg \frac{1 + r}{1 - r} \tag{3.3}$$

este aproximativ repartizată normal de medie

$$\frac{1}{2} \ln \frac{1 + \rho}{1 - \rho}$$

şi dispersie $1/(n-3)$. Astfel dacă este adevărată ipoteza $\rho = 0$, $z\sqrt{n-3}$ este repartizată $N(0,1)$.

Pentru a verifica ipoteza conform căreia două selecţii provin din populaţii cu acelaşi coeficient de corelaţie ρ, se foloseşte statistica.

$$\frac{z_1 - z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

şi tabele cu repartiţia normală.

Scanned with OKEN Scanner

Exemplul 1. (Gh. Zgură, Al. C. Dorin și V. Gh. Vodă: *Studiul deformării plastice cu ajutorul teoriei regresiei și corelației.* „Studii și cercetări de Mecanică Aplicată" 1972).

În urma determinării variației unghiului de revenire elastică β, în funcție de raza relativă de îndoire $\frac{r}{s}$, pentru un unghi de îndoire $\alpha = 30$, s-au obținut următoarele rezultate pentru oțelul OLC 10 — STAS

r/s	0,4	1,7	3	5	5,3	1,2	3,6	10	12	12,5
β	0	1	2	3,5	4	5	6	7	8,5	9

Dintr-o reprezentare grafică sumară rezultă că între β și r/s poate fi stabilită o dependență liniară.

Se verifică această ipoteză pe baza coeficientului de corelație. Calculele corespunzătoare se fac pe baza datelor prezentate sistematic în tabelul 8.1 în care cu x_i s-au notat valorile pentru r/s iar cu y_i valorile determinate pentru β.

Mediile de selecție sînt

$$\bar{x} = \frac{1}{10}\sum_{1}^{10} x_i = 6{,}62, \quad \bar{y} = \frac{1}{10}\sum_{1}^{10} y_i = 4{,}6,$$

iar dispersiile de selecție,

$$S_1 = 4{,}2, \quad S_2 = 3{,}08.$$

Din tabelul 8.1 și din (2.15), obținem pentru coeficientul de corelație de selecție valoarea

$$r_{(x,y)} = \frac{115{,}946}{\sqrt{158{,}4960 \cdot 8{,}49}} = \frac{115{,}946}{\sqrt{116{,}190}} \simeq 1.$$

Se observă o strînsă dependență liniară între (β) și $\frac{r}{S}$. În acest caz ecuația de regresie este

$$\beta = 4{,}6 + \frac{3{,}08}{4{,}20}\left(\frac{r}{S} - 6{,}62\right).$$

Exemplul 2. (D. V. Iliescu. V. Gh. Vodă: *O metodă de reducere a numărului de probe în controlul statistic de recepție, folosind analiza de regresie*, „Revista de Statistică", nr. 10 1972.)

Uneori, în practică, la efectuarea controlului de recepție, extragerea probelor pentru control nu se poate face cu totul

Scanned with OKEN Scanner

Tabelul 8.1

Nr.	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	12,5	9	5,88	4,4	34,5744	19,3600	25,800
2	12	8,5	5,38	3,9	28,9444	15,2100	21,000
3	10	7	3,38	2,4	11,4244	5,7600	8,100
4	8,6	6	1,98	1,4	3,9204	1,9600	2,770
5	7,2	5	0,58	0,4	0,3364	0,1600	0,232
6	5,8	4	—0,82	—0,6	0,6724	0,3600	0,492
7	5	3,5	—1,62	—1,1	2,6244	1,2100	1,782
8	3	2	—0,62	—2,6	13,1044	6,7600	3,400
9	1,7	1	—4,92	—3,6	24,2064	12,9600	17,720
10	0,4	0	—6,22	—4,6	38,6884	21,1600	28,650
Σ	66,2	46,0			158,4960	84,9000	115,946

întîmplător — din motive de ordin tehnologic, economic etc. intervenind un oarecare grad de determinare.

Pentru a clarifica lucrările vom considera o situație ce se întîlnește în industria firelor și fibrelor sintetice (rețelele cord din fire relon ce se folosesc la fabricarea anvelopelor).

Situația este următoarea: un lot livrat de rețele cord este în același timp și un lot de producție el fiind format dintr-un anumit număr de role (de obicei 9—10 role).

Deoarece din fiecare rolă se fabrică un număr destul de mare de anvelope, este necesar ca fiecare din role să corespundă din punct de vedere calitativ.

În această situație, ideal ar fi să se controleze fiecare rolă în parte pentru luarea deciziei de acceptare sau respingere a lotului.

Acest lucru nu este însă posibil din următoarele motive:

a) S-ar conduce la un număr mult prea mare de încercări.

b) Rolele din care se extrag probele pentru control suportă o țesere specială care în mod curent este economic să fie făcută doar la prima și ultima rolă a lotului de producție.

Considerînd pentru exemplificare o caracteristică de calitate cu limitare inferioară (rezistența la rupere a rețelelor cord) vom nota cu:

$\bar{x}_1$ = media măsurătorilor ce compun eșantionul extras din prima rolă,

Scanned with OKEN Scanner

$\bar{x}_2$ = media măsurătorilor ce compun eşantionul extras din ultima rolă,

L_i = limita de control inferioară.

Condiţiile de acceptare a lotului privitor la caracteristica studiată sînt:

$$\bar{x}_1 \geqslant L_i \text{ şi } \bar{x}_2 \geqslant L_i.$$

Dacă se presupun îndeplinite aceste două condiţii se poate aprecia că toate rolele intermediare sînt de calitate corespunzătoare.

Să subliniem faptul că dacă între momentele producerii celor două role pentru control intervine o cădere a procesului de producţie, concluzia de mai sus îşi pierde valabilitatea; de obicei însă procesul nu poate trece de la o stare de cădere înapoi la starea normală fără intervenţia operatorului (deci aceste situaţii nu pot trece neobservate).

Pentru studierea comportării celor două medii de selecţie, vom efectua o analiză de regresie şi corelaţie. În afara concluziilor strict legate de această analiză, ideea constă în determinarea unei valori a variabilei $\bar{x}_1$ — notată $\bar{x}_{1\alpha}$ — pentru care dacă $\bar{x}_1 \geqslant \bar{x}_{1\alpha}$ nu mai este necesară operaţia de control pentru ultima rolă, considerîndu-se cu nivelul de semnificaţie α ales (care de fapt este riscul luării acestei decizii) că $\bar{x}_2 > L_i$, decizia fiind de acceptare a lotului numai pe baza rezultatelor controlului primei role.

Acest lucru se poate observa din figura 8.2, unde în mod convenţional s-a reprezentat dreapta de regresie cu zona ei de încredere.

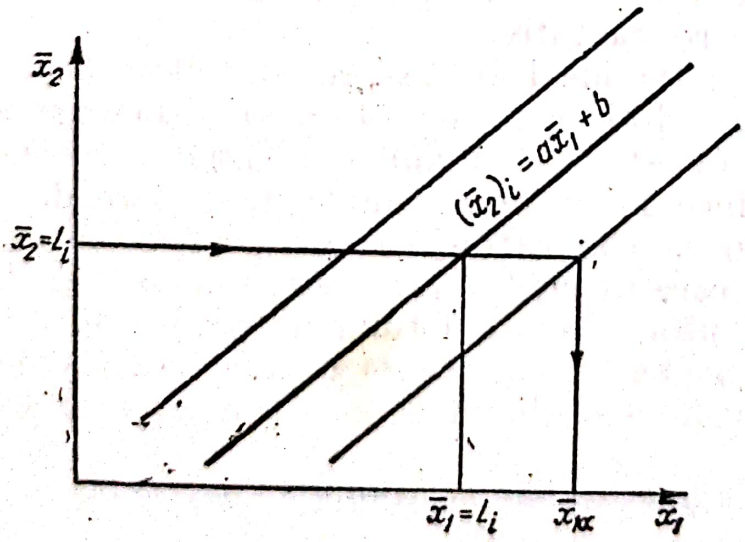


Fig. 8.2

Din figura 8.2 se observă că pentru toți $\bar{x}_1 > \bar{x}_{1\alpha}$, rezultă cu probabilitatea $(1 - \alpha)$ că $\bar{x}_2 > L_t$, deci:

1) dacă $\bar{x}_1 < L_t$ se refuză lotul fără a se mai controla cea de-a doua rolă,

2) dacă $L_t \leqslant \bar{x}_1 \leqslant \bar{x}_{1\alpha}$ se controlează și a doua rolă, decizia de acceptare luîndu-se dacă $\bar{x}_2 \geqslant L_t$,

3) dacă $\bar{x}_1 > \bar{x}_{1\alpha}$ se acceptă lotul fără a se mai controla cea de-a doua rolă.

După cum s-a mai spus, din reprezentarea grafică a perechilor de valori $(\bar{x}_{1i}, \bar{x}_{2i})$ unde $i = 1, 2, ..., m$ — în cazul nostru $m = 158$ s-a dedus că este vorba de o regresie liniară.

Fie $\bar{x}_1$ și $\bar{x}_2$ cele două variabile aleatoare care reprezintă mediile de selecție pentru caracteristica „rezistență la rupere" a rețelelor cord referitoare la prima și ultima rolă a lotului de producție.

Aceste variabile sînt considerate niște variabile aleatoare normale

$$N\left(\mu_1, \frac{\sigma}{\sqrt{n}}\right), \quad N\left(\mu_2, \frac{\sigma}{\sqrt{n}}\right).$$

Valorile de selecție ale acestor variabile $\bar{x}_{1i}$ și $\bar{x}_{2i}$ au fost calculate pe baza a douăzeci de măsurători fiecare (au fost în total prelucrate 6 320 de măsurători individuale).

Dependența dintre $\bar{x}_2$ și $\bar{x}_1$ este de forma

$$(\bar{x}_2)_{calculat} = a\bar{x}_1 + b.$$

Coeficienții a și b ai formei liniare au fost calculați folosind metoda celor mai mici pătrate, obținîndu-se următoarele valori:

$$a = \frac{n\sum_1^n \bar{x}_{1i}\bar{x}_{2i} - \left(\sum_1^n \bar{x}_{1i}\right)\left(\sum_1^n \bar{x}_{2i}\right)}{n\sum_1^n (\bar{x}_{1i})^2 - \left(\sum_1^n \bar{x}_{1i}\right)^2} = 0{,}838,$$

$$b = \frac{\left(\sum_1^n \bar{x}_{1i}^2\right)\left(\sum_1^n \bar{x}_{2i}\right) - \left(\sum_1^n \bar{x}_{1i}\bar{x}_{2i}\right)\left(\sum_1^n \bar{x}_{1i}\right)}{n\sum_1^n (\bar{x}_{1i})^2 - \left(\sum_1^n \bar{x}_{1i}\right)^2} = 3{,}383.$$

Scanned with OKEN Scanner

Avem deci

$$(\bar{x}_2)_c = 0{,}838\bar{x}_1 + 3{,}383.$$

Intensitatea legăturii liniare dintre valorile lui $\bar{x}_2$ și $\bar{x}_1$ a fost măsurată cu ajutorul coeficientului de corelație de selecție, dat de formula

$$r(\bar{x}_1, \bar{x}_2) = \frac{n\sum_1^n \bar{x}_{1i}\bar{x}_{2i} - \left(\sum_1^n \bar{x}_{1i}\right)\left(\sum_1^n x_{2i}\right)}{\sqrt{\left[n\sum_1^n x_{1i}^2 - \left(\sum_1^n \bar{x}_{1i}\right)^2\right]\left[n\sum_1^n \bar{x}_{2i}^2 - \left(\sum_1^n \bar{x}_{2i}\right)^2\right]}} = 0{,}83$$

fapt care arată că această legătură liniară este suficient de puternică.

Aplicînd transformarea lui Fisher

$$z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

vom verifica ipoteza $H_0 : \rho = 0{,}9$ care exprimă faptul că în realitate coeficientul de corelație teoretic este mai mare decît cel de selecție (0,83) și anume 0,9.

Avem

$$z = \frac{1}{2}\ln\frac{1+0{,}83}{1-0{,}83} = \frac{1}{2}\ln\frac{183}{17} = \frac{1}{2}\ln 10{,}76 \simeq \frac{1}{2}\cdot 2{,}3 \lg 10{,}76 =$$

$$= 1{,}15 \cdot 1{,}031,$$

$$z' = \frac{1}{2}\ln\frac{1+0{,}9}{1-0{,}9} = \frac{1}{2}\ln 19 \simeq \frac{1}{2}\cdot 2{,}3 \lg 19 = 1{,}15 \cdot 1{,}278,$$

$$\sqrt{n-3} = \sqrt{158-3} = \sqrt{155} \simeq 12{,}45.$$

Calculăm statistica lui Fisher:

$$F(\rho) = |\sqrt{n-3}(z-z')| = |12{,}45(1{,}15 \cdot 1{,}031 - 1{,}15 \cdot 1{,}278)| =$$

$$= 12{,}45 \cdot 0{,}247 = 3{,}07515.$$

Scanned with OKEN Scanner

Dacă $F(\rho) < \tau_\varepsilon$ atunci se acceptă H_0 cu nivelul de semnificație ε. τ_ε este dat de ecuația

$$\frac{1}{\sqrt{2\pi}} \int_{-\tau_\varepsilon}^{\tau_\varepsilon} e^{-\frac{t^2}{2}} dt = 1 - \varepsilon.$$

Pentru $\varepsilon = 0{,}05$ găsim $\tau_\varepsilon = 1{,}96$.

Deoarece $F(\rho) = 3{,}07515 > 1{,}96$ ipoteza H_0 se respinge.

Vom da în această situație un interval de încredere pentru coeficientul de corelație teoretic.

Se știe de asemenea că un interval de încredere pentru coeficientul de corelație teoretic este dat de

$$I_\rho = \left[\frac{e^{2A} - 1}{e^{2A} + 1}, \frac{e^{2B} - 1}{e^{2B} + 1}\right]$$

unde

$$A = z - \frac{\tau_\varepsilon}{\sqrt{n-3}} \quad \text{și} \quad B = z + \frac{\tau_\varepsilon}{\sqrt{n-3}}$$

(deci cu probabilitate $1 - \varepsilon$ coeficientul de corelație teoretic se găsește în intervalul I_ρ).

Avem

$$A = z - \frac{\tau_\varepsilon}{\sqrt{n-3}} = 1{,}15 \cdot 1{,}031 + \frac{1{,}96}{12{,}45} =$$

$$= 1{,}18565 + 0{,}157 = 1{,}34265,$$

$$B = z + \frac{\tau_\varepsilon}{\sqrt{n-3}} = 1{,}18565 - 0{,}157 = 1{,}02865.$$

Ținînd cont că $e \simeq 2{,}71$ și efectuînd calculele indicate de expresia lui I_ρ găsim în final

$$I_\rho = [0{,}802,\ 0{,}864].$$

Scanned with OKEN Scanner

CAPITOLUL
9

REGRESIE MULTIPLĂ ŞI ANALIZĂ DISPERSIONALĂ

§ 1. Introducere

De obicei, în procesul de cercetare a legăturilor cauzale dintre fenomenele de masă avem de-a face cu contopirea acţiunii unei multitudini de cauze, care determină variaţia fenomenului efect. În astfel de cazuri, calculul corelaţiei nu se poate limita la variabile perechi (cap. 7), ci trebuie să includă şi alte variabile independente cu influenţă semnificativă asupra variabilei dependente studiate. În acest capitol ne vom ocupa de regresia lui y asupra a r variabile independente x_i.

§ 2. Regresia lui y asupra a două variabile independente. Estimarea parametrilor

Presupunem că y este repartizată normal de medie şi dispersie σ^2 şi că observaţiile sînt independente. De asemenea presupunem că μ este o funcţie liniară de două variabile independente x_1 şi x_2,

$$\mu = \alpha + \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2). \tag{2.1}$$

Vom cere, în cele ce urmează, cunoaşterea valorilor x_{1k}, x_{2k} pentru fiecare valoare a lui k.

Presupunem că avem n triplete de observaţie (y_k, x_{1k}, x_{2k}) şi ne propunem să determinăm estimaţiile a, b_1, b_2 şi s^2 pentru parametrii α, β_1, β_2 şi σ^2.

Ecuaţia estimatoare este

$$Y = a + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2), \tag{2.2}$$

Scanned with OKEN Scanner

iar suma pătratelor abaterilor dintre valorile observate y_k și valorile date de ecuația estimatoare este

$$S = \sum_1^n (y_k - Y_k)^2 = \sum_1^n \{y_k -$$

$$- [a + b_1(x_{1k} - \bar{x}_1) + b_2(x_{2k} - \bar{x}_2)]\}^2. \tag{2.3}$$

Pentru a minimiza (2.3) considerăm derivatele în raport cu a, b_1 și b_2 și le egalăm cu zero.

$$\frac{\partial S}{\partial a} = -2\sum_1^n \{y_k - [a + b_1(x_{1k} - \bar{x}_1) + b_2(x_{2k} - \bar{x}_2)]\} = 0, \tag{2.4}$$

$$\frac{\partial S}{\partial b_1} = -2\sum_1^n \{y_k - [a + b_1(x_{1k} - \bar{x}_1) + b_2(x_{2k} - \bar{x}_2)]\}(x_{1k} - \bar{x}_1) = 0, \tag{2.5}$$

$$\frac{\partial S}{\partial b_2} = -2\sum_1^n \{y_k - [a + b_1(x_{1k} - \bar{x}_1) + b_2(x_{2k} - \bar{x}_2)]\}(x_{2k} - \bar{x}) = 0. \tag{2.6}$$

Deoarece

$$\sum_1^n (x_{1k} - \bar{x}_1) = \sum_1^n (x_{2k} - \bar{x}_2) = 0,$$

din (2.4) obținem

$$\sum_1^n y_k - na - b_1\sum_1^n (x_{1k} - \bar{x}_1) - b_2\sum_1^n (x_{2k} - \bar{x}_2) =$$

$$= \sum_1^n y_k - na = 0.$$

de unde

$$a = \frac{1}{n}\sum_1^n y_k. \tag{2.7}$$

Scanned with OKEN Scanner

Ecuaţia (2.4) conduce la

$$\sum_1^n (y_k - Y_k) = 0, \tag{2.8}$$

iar ecuaţiile (2.5) şi (2.6) conduc la

$$\sum_1^n (y_k - Y_k)(x_{1k} - \bar{x}_1) = \sum_1^n (y_k - Y_k)(x_{2k} - \bar{x}_2) = 0 \tag{2.9}$$

şi

$$\begin{cases} b_1 \sum_1^n (x_{1k} - \bar{x}_1)^2 + b_2 \sum_1^n (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) = \\ \qquad = \sum_1^n y_k (x_{1k} - \bar{x}_1) \\ b_1 \sum_1^n (x_{1k} - \bar{x}_1)^2 (x_{2k} - \bar{x}_2) + b_2 \sum_1^n (x_{2k} - \bar{x}_2)^2 = \\ \qquad = \sum_1^n y_k (x_{2k} - \bar{x}_2) \end{cases} \tag{2.10}$$

Ecuaţiile sistemului (2.10) sînt cunoscute sub numele de *ecuaţii normale* şi deoarece conţin sume de pătrate şi sume de produse ce pot fi evaluate din observaţie, soluţiile b_1 şi b_2 se pot obţine.

Suma pătratelor pentru s^2 poate fi obţinută din (2.3) prin înlocuirea lui a, b_1 şi b_2 cu soluţiile obţinute [(2.7) şi din sistemul (2.10)].

Observaţii. Putem scrie

$$y_k - \bar{y} = (y_k - Y_k) + (Y_k - \bar{y}),$$

de unde ridicînd la pătrat ambii membri şi ţinînd seama că

$$2 \sum_1^n (y_k - Y_k)(Y_k - \bar{y}) = 2b_1 \sum_1^n (y_k - Y_k)(x_{1k} - \bar{x}_1) +$$

$$+ 2b_2 \sum_1^n (y_k - Y_k)(x_{2k} - \bar{x}_2) = 0,$$

Scanned with OKEN Scanner

obținem

$$(2.11)\qquad \sum_1^n (y_k - \bar{y})^2 = \sum_1^n (y_k - Y_k)^2 + \sum_1^n (Y_k - \bar{y})^2$$

sau

$$(2.12)\qquad \sum_1^n (y_k - Y_k)^2 = \sum_1^n (y_k - \bar{y})^2 - \sum_1^n (Y_k - \bar{y})^2.$$

Înmulțind ecuațiile normale (2.10) prin b_1 respectiv b_2 obținem

$$b_1^2 \sum_1^n (x_{1k} - \bar{x}_1)^2 + b_1 b_2 \sum_1^n (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) =$$

$$= b_1 \sum_1^n y_k (x_{1k} - \bar{x}_1),$$

$$b_1 b_2 \sum_1^n (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) + b_2 \sum_1^n (x_{2k} - \bar{x}_2)^2 =$$

$$= b_2 \sum_1^n y_k (x_{2k} - \bar{x}_1),$$

de unde

$$(2.13)\qquad b_1 \sum_1^n y_k (x_{1k} - \bar{x}_1) + b_2 \sum_1^n y_k (x_{2k} - \bar{x}_1) =$$

$$= \sum_1^n [b_1^2 (x_{1k} - \bar{x}_1)^2 + 2 b_1 b_2 (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) + b_2^2 (x_{2k} - \bar{x}_2)^2].$$

sau

$$b_1 \sum_1^n y_k (x_{1k} - \bar{x}_1) + b_2 \sum_1^n y_k (x_{2k} - \bar{x}_1) =$$

$$(2.14)\qquad = \sum_1^n [b_1 (x_{1k} - \bar{x}_1) + b_2 (x_{2k} - \bar{x}_2)]^2 = \sum_1^n (Y_k - \bar{y})^2.$$

Din (2.12) și (2.14) se poate determina suma pătratelor care apare în s^2. Deoarece parametrii a, b_1 și b_2 au fost determinați cu ajutorul observațiilor, s^2 are $n-3$ grade de libertate.

Scanned with OKEN Scanner

§ 3. Repartiţii în cadrul modelului

Mai întîi vom rezolva sistemul (2.10). În acest scop vom înmulţi prima ecuaţie prin $\sum_1^n (x_{2k} - \bar{x}_2)^2$, a doua prin $\sum_1^n (x_{1k} - x_1)^2$ iar apoi le vom scade. Obţinem

$$b_1 = \frac{(\Delta\delta_1)\delta_2^2 - (\Delta\delta_2)\delta_{12}}{\delta_1^2\delta_2^2 - \delta_{12}^2} \tag{3.1}$$

unde

$$\delta_1^2 = \sum_1^n (x_{1k} - \bar{x}_1)^2, \qquad \delta_2^2 = \sum_1^n (x_{2k} - \bar{x}_2)^2$$

$$\delta_{12} = \sum_1^n (x_{1k} - \bar{x}_1)\,(x_{2k} - \bar{x}_2), \tag{3.2}$$

$$\Delta\delta_1 = \sum_1^n y_k(x_{1k} - \bar{x}_1), \qquad \Delta\delta_2 = \sum_1^n y_k(x_{2k} - \bar{x}_2)$$

(3.1) se mai poate pune sub forma

$$b_1 = \frac{1}{\delta_1^2\delta_2^2 - \delta_{12}^2}\left\{\sum_1^n [\delta_1^2(x_{1k} - \bar{x}_1) - \delta_{12}(x_{2k} - \bar{x}_2)]\, y_k\right\}, \tag{3.3}$$

ceea ce arată că b_1 este o funcţie liniară de y şi cum y este repartizat normal, b_1 va fi de asemenea repartizat normal.

Dispersia lui b_1 este dată de

$$D^2(b_1) = \frac{1}{(\delta_1^2\delta_2^2 - \delta_{12}^2)^2}\sum_1^n [\delta_1^2(x_{1k} - \bar{x}_1) - \delta_{12}(x_{2k} - \bar{x}_2)]^2\, D^2(y_k) =$$

$$= \frac{\delta_2^2}{\delta_1^2\delta_2^2 - \delta_{12}^2}\sigma^2, \tag{3.4}$$

şi analog

$$D^2(b_2) = \frac{\delta_1^2}{\delta_1^2\delta_2^2 - \delta_{12}^2}\sigma^2, \tag{3.5}$$

$$\operatorname{cov}(b_1, b_2) = -\frac{\delta_{12}}{\delta_1^2\delta_2^2 - \delta_{12}^2}\sigma^2, \tag{3.6}$$

Scanned with OKEN Scanner

de unde

$$\rho_{b_1,\,b_2} = -\frac{\delta_{12}}{\sqrt{\delta_1^2\delta_2^2}}. \tag{3.7}$$

Dacă în (3.4) şi (3.5) înlocuim σ^2 prin s^2, relaţiile astfel obţinute pot fi folosite pentru verificarea ipotezelor $\beta_1 = 0$ şi $\beta_2 = 0$. De la repartiţia normală bidimensională se ştie că

$$\frac{1}{1-\rho^2_{b_1,\,b_2}}\left[\left(\frac{b_1-\beta_1}{\sqrt{D^2(b_1)}}\right)^2 - 2\rho_{b_1,\,b_2}\left(\frac{b_1-\beta_1}{\sqrt{D^2(b_1)}}\right)\left(\frac{b_2-\beta_2}{\sqrt{D^2(b_2)}}\right) + \left(\frac{b_2-\beta_2}{\sqrt{D^2(b_2)}}\right)^2\right], \tag{3.8}$$

este repartizată χ^2 cu 2 grade de libertate.

Din (3.7)

$$1 - \rho^2_{b_1,\,b_2} = 1 - \frac{\delta_{12}^2}{\delta_1^2\delta_2^2} = \frac{\delta_1^2\delta_2^2 - \delta_{12}^2}{\delta_1^2\delta_2^2},$$

ceea ce permite să scriem (3.8), după ce-l înmulţim cu σ^2, sub forma

$$(b_1-\beta_1)^2\,\delta_1^2 + 2(b_1-\beta_1)(b_2-\beta_2)\,\delta_{12} + (b_2-\beta_2)^2\,\delta_2^2 \tag{3.9}$$

cantitate care este repartizată ca $\sigma^2\chi^2$ (2).

§ 4. Studiul regresiei cu ajutorul analizei dispersionale

Ţinînd seama de (2.1) şi (2.2), diferenţa dintre observaţiile y_k şi valorile adevărate η_k poate fi scrisă

$$\begin{aligned} y_k - \eta_k &= (y_k - Y_k) + (Y_k - \eta_k) = \\ &= (y_k - Y_k) + (a-\alpha) + (b_1-\beta_1)(x_{1k} - \bar{x}_1) + \\ &\quad + (b_2-\beta_2)(x_{2k} - \bar{x}_2). \end{aligned} \tag{4.1}$$

Ridicînd la pătrat, sumînd după k de la 1 la n şi ţinînd seama de relaţiile (2.8) şi (2.9) obţinem

$$\begin{aligned} \sum_1^n (y_k - \eta_k)^2 &= \sum_1^n (y_k - Y_k)^2 + n(a-\alpha)^2 + \\ &+ (b_1-\beta_1)^2 \sum_1^n (x_{1k} - \bar{x}_1)^2 + (b_2-\beta_2)\sum_1^n (x_{2k} - \bar{x}_2)^2 + \\ &+ 2(b_1-\beta_1)(b_2-\beta_2)\sum_1^n (x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2). \end{aligned} \tag{4.2}$$

Scanned with OKEN Scanner

Membrul stîng al relației (4.2) este repartizat ca $\sigma^2\chi^2(n)$. În membrul drept $\sum_1^n (y_k - Y_k)^2$ are $(n-3)$ grade de libertate, deoarece $(y_k - Y_k)$ satisfac condițiile (2.8) și (2.9) $n(a-\alpha)^2$ are un grad de libertate și cum

$$\frac{a-\alpha}{\sigma/\sqrt{n}} \sim N(0,1)$$

urmează că

$$n(a-\alpha)^2 \sim \sigma^2\chi^2(1).$$

(3.9) ne spune că suma ultimilor trei termeni este repartizată ca $\sigma^2\chi^2(2)$. Astfel condițiile teoremei lui Cochran sînt satisfăcute și

$$\sum_1^n (y_k - Y_k)^2, \quad n(a-\alpha)^2 \text{ și } (3.9)$$

sînt independente și repartizate ca $\sigma^2\chi^2$.

Deoarece

$$s^2 = \frac{1}{n-3}\sum_1^n (y_k - Y_k)^2, \tag{4.3}$$

$$M(s^2) = \frac{1}{n-3} M[\sigma^2\chi^2(n-3)] = \sigma^2. \tag{4.4}$$

Pentru a verifica ipoteza $\beta_1 = \beta_2 = 0$ procedăm după cum urmează.

Deoarece (3.9) este repartizat ca $\sigma^2\chi^2(2)$,

$$M\left[\frac{1}{2}(b_1-\beta_1)^2\delta_1^2 + 2(b_1-\beta_1)(b_2-\beta_2)\delta_{12} + \right.$$

$$\left. + (b_2-\beta_2)^2\delta_2^2\right] = \sigma^2 \tag{4.5}$$

ceea ce ne arată că (3.9) împărțit prin 2 este o estimație pentru σ^2, estimație independentă de s^2, care este la rîndu-i o estimație pentru σ^2. Urmează că raportul lor va fi repartizat $F(2, n-3)$, ceea ce ne permite să afirmăm că dacă ipoteza nulă $\beta_1 = \beta_2 = 0$ este adevărată, atunci

$$\frac{1}{2s^2}(b_1^2\delta_1^2 + 2b_1b_2\delta_{12} + b_2^2\delta_2^2) \sim F(2,n-3). \tag{4.6}$$

Scanned with OKEN Scanner

În (4.6) în loc de cantitatea din paranteză, ținînd seama de (2.13), folosim

$$b_1 \sum_1^n y_k(x_{1k} - \bar{x}_1) + b_2 \sum_1^n y_k(x_{2k} - \bar{x}_2).$$

Deoarece

$$\eta = M(Y) = \alpha + \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2)$$

și deoarece a este independentă de b_1 și b_2

$$D^2(Y) = D^2(a) + (x_1 - \bar{x}_1)^2 D^2(b_1) + (x_2 - \bar{x}_1)^2 D^2(b^2) +$$

$$+ 2(x_1 - \bar{x}_1)(x_2 - \bar{x}_2) \operatorname{cov}(b_1, b_2) =$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x_1 - \bar{x}_1)^2 \delta_2^2 - 2(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)\delta_{12} + (x_2 - \bar{x}_2)^2 \delta_1^2}{\delta_1^2 \delta_2^2 - \delta_{12}^2} \right].$$

Ecuația (2.11) stă la baza tabelului 10.1 și din (2.13) suma pătratelor date de regresie este

$$\sum_1^n (Y_k - \bar{y})^2 = b_1 \Delta\delta_1 + b_2 \Delta\delta_2. \tag{4.8}$$

Tabelul 10. 1

Abaterile	Suma pătratelor	Nr. gradelor de libertate	2/3
1	2	3	4
date de regresie	$b_1 \Delta\,\delta_1 + b_2 \Delta\,\delta_2$	2	s_2^2
față de planul de regresie	$\sum_1^n (y_k - Y_k)^2$	$n - 3$	s^2
Total (prin diferență)	$\sum_1^n (y_k - \bar{y})^2$	$n - 1$	

Scanned with OKEN Scanner

§ 5. Regresia în cazul mai multor variabile independente

Presupunem că avem punctele

$$(y_k,\ x_{1k},\ ...,\ x_{rk}),\quad k = 1, 2, ..., n$$

şi că y este repartizată $N(\eta, \sigma^2)$ unde η este funcţia liniară

$$\eta = \alpha + \beta_1(x_1 - \bar{x}_1) + \beta_2(x_2 - \bar{x}_2) + ... + \beta_r(x_r - \bar{x}_r). \tag{5.1}$$

Ecuaţia estimatoare este

$$Y = a + b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + ... + b_r(x_r - \bar{x}_r) \tag{5.2}$$

şi estimaţiile $a, b_1, ..., b_r$ sînt obţinute minimizînd suma pătratelor abaterilor dintre valorile observate şi cele date de ecuaţia estimatoare.

$$S = \sum_1^n (y_k - Y_k)^2 = \sum_1^n \{y_k - [a + b_1\ (x_{1k} - \bar{x}_1) + ... + b_r(x_{rk} - \bar{x}_r)]\}^2, \tag{5.3}$$

prin derivare în raport cu $a, b_1, ..., b_r$ şi egalînd cu zero. Astfel obţinem

$$a = \frac{1}{n}\sum_1^n y_k \tag{5.4}$$

şi sistemul de ecuaţii normale:

$$\begin{aligned} &b_1\delta_1^2 + b_2\delta_{12} + ... + b_r\delta_{1r} = \Delta\delta_1,\\ &b_1\delta_{12} + b_2\delta_2^2 + ... + b_r\delta_{2r} = \Delta\delta_2,\\ &\vdots\\ &b_1\delta_{1r} + b_2\delta_{2r} + ... + b_r\delta_r^2 = \Delta\delta_r \end{aligned} \tag{5.5}$$

unde

$$\delta_i^2 = \sum_{k=1}^n (x_{ik} - \bar{x}_i)^2,\quad \delta_{ij} = \sum_1^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j),$$

$$\Delta\delta_i = \sum_{k=1}^n y_k(x_{ik} - \bar{x}_i),\quad 1 \leqslant i,\ j \leqslant r. \tag{5.6}$$

Scanned with OKEN Scanner

Sistemul (5.5) se poate rezolva în raport cu $b_1, \ldots, b_r$. Obţinem că b_i este o funcţie lineară de y_i şi cum y_i este repartizată normal, urmează că şi b_i este repartizat normal.

§ 6. Analiza dispersională în cazul mai multor variabile independente

Putem scrie

$$y_k - \eta_k = (y_k - Y_k) + (a - \alpha) + (b_1 - \beta_1)(x_{1k} - \bar{x}_1) + \tag{6.1}$$
$$+ \ldots (b_r - \beta_r)(x_{rk} - \bar{x}_r),$$

de unde ridicînd la pătrat şi sumînd

$$\sum_1^n (y_k - \eta_k)^2 = \sum_1^n (y_k - Y_k)^2 + n(a - \alpha)^2 +$$

$$+ \sum_1^r (b_i - \beta_i)^2 \delta_1^2 + 2 \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (b_i - \beta_i)(b_j - \beta_j)\delta_{ij}. \tag{6.2}$$

Membrul stîng al relaţiei (6.2) este repartizat ca $\sigma^2\chi^2(n)$. În membrul drept

$\sum_1^n (y_k - Y_k)^2$ este repartizată ca $\sigma^2\chi^2(n - r - 1)$,

$n(a - \alpha)^2$ este repartizată ca $\sigma^2\chi^2(1)$,

iar

$$\sum_1^n (b_i - \beta_i)^2 \delta_1^2 + 2 \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (b_i - \beta_i)(b_j - \beta_j)\delta_{ij}. \tag{6.3}$$

ca $\sigma^2\chi^2(r)$, fiecare dintre aceste componente fiind independentă de oricare alta.

Dacă punem

$$s^2 = \frac{1}{n - r - 1} \sum_1^n (y_k - Y_k)^2, \tag{6.4}$$

atunci $M(s^2) = \sigma^2$.

a este repartizată $N\left(\alpha, \frac{\sigma^2}{n}\right)$.

Scanned with OKEN Scanner

Avem

$$D^2(Y) = D^2(a) + \sum_1^r (x_i - \bar{x})^2 D^2(b_i) +$$

$$+ 2\sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (x_i - \bar{x}_i)(x_j - \bar{x}_j) \operatorname{cov}(b_i, b_j). \tag{6.5}$$

Ipoteza conform căreia $\beta_i = 0$, pentru toţi i se verifică după cum urmează.

Deoarece (6.3) este repartizată ca $\sigma^2\chi^2(r)$,

$$M\left\{\frac{1}{r}\left[\sum_1^n (b_i - \beta_i)^2 \delta_i^2 + \right.\right.$$

$$\left.\left. + 2\sum_{i=1}^{r-1} \sum_{j=i+1}^{r} (b_i - \beta_i)(b_j - \beta_j)\delta_{ij}\right]\right\} = \sigma^2 \tag{6.6}$$

şi este independentă de s^2 definită în (6.4). Dacă este adevărată ipoteza nulă $\beta_i = 0$ pentru toţi i

$$\frac{1}{rs^2}\left[\sum_1^n b_i^2\delta_i^2 + 2\sum_{i=1}^{r-1}\sum_{j=i+1}^{r} b_i b_j \delta_{ij}\right] \sim F(r, n-r-1). \tag{6.7}$$

§ 7. Coeficientul de corelaţie parţial

Ca şi în cazul repartiţiei unui vector aleator bidimensional (x_1, x_2) normal, cînd integrînd densitatea de repartiţie în raport cu x_2 obţinem densitatea de repartiţie marginală a lui x_1, care era densitatea corespunzătoare unei repartiţii normale unidimensională, în cazul repartiţiei unui vector (x_1, x_2, x_3) normal, cînd integrăm densitatea de repartiţie în raport cu x_2 obţinem densitatea de repartiţie corespunzătoare unei repartiţii normale bidimensionale în x_1 şi x_3. Această repartiţie normală bidimensională în x_1 şi x_3 va avea proprietăţile obişnuite ale repartiţiei normale bidimensionale.

Scanned with OKEN Scanner

Drepata de regresie a lui x_1 asupra lui x_3 este

$$M(x_1 \mid x_3) = \mu_1 + \beta_{13}(x_3 - \mu_3). \tag{7.1}$$

Analog, dacă integrăm densitatea de repartiție a vectorului normal (x_1, x_2, x_3) în raport cu x_1, obținem densitatea de repartiție a vectorului (x_2, x_3) care este de asemenea normală.

Dreapta de regresie a lui x_2 asupra lui x_3 este

$$M(x_2 \mid x_3) = \mu_2 + \beta_{23}(x_3 - \mu_3). \tag{7.2}$$

Fie

$$x_{1\cdot 3} = x_1 - M(x_1 \mid x_3) = (x_1 - \mu_1) - \beta_{13}(x_3 - \mu_3) \tag{7.3}$$

$$x_{2\cdot 3} = x_2 - M(x_2 \mid x_3) = (x_2 - \mu_2) - \beta_{23}(x_3 - \mu_3). \tag{7.4}$$

Coeficientul de corelație parțial dintre x_1 și x_2, eliminînd x_3 este

$$\rho_{12\cdot 3} = \frac{\operatorname{cov}(x_{1\cdot 3}, x_{2\cdot 3})}{D(x_{1\cdot 3})\, D(x_{2\cdot 3})}. \tag{7.5}$$

Vom găsi o expresie a acestui coeficient în funcție de ρ_{12}, ρ_{23}, ρ_{13}. Din (7.3) și din (8.2.12) relația $\beta_{13} = \rho_{13}\,\sigma_1/\sigma_3$, obținem

$$D^2(x_{1\cdot 3}) = D^2(x_1) + \beta_{13}^2\, D^2(x_3) + 2(-\beta_{13}) \operatorname{cov}(x_1, x_3) = \tag{7.6}$$

$$= \sigma_1^2 + \left(\rho_{13}\frac{\sigma_1}{\sigma_3}\right)^2 \sigma_3^2 - 2\left(\rho_{13}\frac{\sigma_1}{\sigma_2}\right) \rho_{13}\sigma_1\sigma_3 = \sigma_1^2(1 - \rho_{13}^2).$$

Analog, obținem

$$D^2(x_{2\cdot 3}) = \sigma_2^2(1 - \rho_{23}^2). \tag{7.7}$$

De asemenea

$$\operatorname{cov}(x_{1\cdot 3}, x_{2\cdot 3}) = M(x_{1\cdot 3}, x_{2\cdot 3}) - M(x_{1\cdot 3})\, M(x_{2\cdot 3}) = \tag{7.8}$$

$$= \sigma_1\sigma_2(\rho_{12} - \rho_{13}\rho_{23}).$$

Ținînd seama în (7.5) de (7.6), (7.7) și (7.8), obținem

$$\rho_{12\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}} \tag{7.9}$$

Scanned with OKEN Scanner

Coeficientul de corelație de selecție parțial $r_{12 \cdot 3}$ are o expresie cu totul analoagă

(7.10) $$r_{12 \cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}) (1 - r_{23}^2)}}$$

Ca și în (8.4.2.) cantitatea

(7.11) $$\sqrt{n-3} \frac{r_{12 \cdot 3}}{\sqrt{1 - r_{12 \cdot 3}^2}} \sim t(n-3)$$

dacă $\rho_{12 \cdot 3} = 0$, fapt ce permite să verificăm ipoteza nulă, conform căreia $\rho_{12.3} = 0$.

Scanned with OKEN Scanner

CAPITOLUL
10

TESTUL SECVENȚIAL

§ 1. Introducere

Toate testele considerate pînă acum au fost bazate pe selecții care aveau un volum bine determinat. Există însă experiențe care pot fi conduse pe baza acumulării unor informații obținute în timpul selecției și corespunzător există metode care ne permit să tragem concluzii asupra caracteristicilor studiate. Aceste metode, datorite lui A. Wald, sînt cunoscute sub numele de metode secvențiale deoarece se operează asupra termenilor succesivi ai șirului de observații pe măsură ce ei sînt obținuți.

§ 2. Testul secvențial al raportului probabilităților (T.S.R.P.)

Să presupunem că vrem să verificăm ipoteza:

H_0 — populația are densitatea de repartiție $f_0(x)$ față de alternativa

H_1 — populația are densitatea de repartiție $f_1(x)$, fără a fixa dinainte volumul selecției.

Analiza secvențială dă un procedeu de a lua una din următoarele trei decizii la fiecare etapă a experimentului.

I. Dacă

$$A < \frac{\prod_1^n f_1(x_i)}{\prod_1^n f_0(x_i)} < B \tag{2.1}$$

facem observația de ordin $(n+1)$, adică continuăm experimentul.

Scanned with OKEN Scanner

II. Dacă

$$\frac{\prod_1^n f_1(x_i)}{\prod_1^n f_0(x_i)} \leqslant A \tag{2.2}$$

experimentul se opreşte şi se acceptă ipoteza H_0.

III. Dacă

$$B \leqslant \frac{\prod_1^n f_1(x_i)}{\prod_1^n f_0(x_i)}.$$

experimentul se opreşte şi se respinge ipoteza H_0 (se acceptă H_1).

A şi B se determină astfel încît probabilităţile erorilor de ordinul I şi II să fie α şi respectiv β.

Deoarece valorile A şi B se obţin cu dificultate, se folosesc aproximaţiile

$$A \simeq \frac{\beta}{1-\alpha}, \quad B \simeq \frac{1-\beta}{\alpha}.$$

§ 3. Metoda lui Girshick, M.A.

Fie X şi Y două variabile aleatoare independente ale căror repartiţii au aceeaşi formă funcţională, presupusă cunoscută, dar conţinînd parametrii θ şi ω necunoscuţi.

Trebuie verificată ipoteza:

$$H : \theta \leqslant \omega \text{ faţă de alternativa } H' : \theta > \omega$$

În acest scop M. A. Girshick (1946) a propus următorul procedeu: se aleg două valori θ_0 şi ω_0 ($\theta_0 < \omega_0$) şi fie H_0 şi H_1 ipotezele:

$$H_0 : \theta = \theta_0; \ \omega = \omega_0, \quad H_1 : \theta = \omega_0; \ \omega = \theta_0$$

În acest caz, a accepta ipoteza H înseamnă a accepta ipoteza H_0, iar a respinge H este echivalent cu a accepta H_1.

Scanned with OKEN Scanner

Atunci, testul secvențial pentru H va fi un T.S.R.P. pentru verificarea ipotezei H_0 față de alternativa H_1.

Dacă notăm

$$f_0(x, y) = f(x; \theta_0)\, f(y; \omega_0) \text{ și } f_1(x,y) = f(x; \omega_0)\, f(y; \theta_0)$$

atunci funcțiile de verosimilitate asociate perechilor de măsurători $(x_1, y_1), \ldots, (x_n, y_n)$ în cazul verificării ipotezelor H_0 și H_1 sînt

$$p_{0n} = \prod_1^n f_0(x_i, y_i) \quad \text{și} \quad p_{1n} = \prod_1^n f_1(x_i, y_i).$$

Construcția T.S.R.P. urmează metoda din paragraful anterior (metoda lui Wald) pe baza raportului $r_n = p_{1n}/p_{0n}$.

Girshick a arătat că reducerea verificării ipotezei H la verificarea ipotezei H_0 se poate efectua numai dacă există o funcție $v(\theta; \omega)$ care satisface relațiile:

1) $v(\theta; \omega) = 0$ dacă $\theta = \omega$,

2) $v(\theta; \omega) < 0$ dacă $\theta < \omega$,

3) $v(\theta; \omega) = -v(\omega; \theta)$.

Această funcție se interpretează ca măsura diferenței dintre θ și ω. Atunci:

a) dacă $v(\theta; \delta) \geqslant \delta$ cu $\delta > 0$, acceptarea lui H este o eroare,

b) dacă $v(\theta; \omega) \leqslant -\delta$, respingerea lui H este o eroare.

Valorile θ_0 și ω_0 se pot alege din condiția $v(\theta_0, \omega_0) = -\delta$.

§ 4. Aplicații ale testelor secvențiale

Se știe că în procesul de producție, un anumit produs se poate obține folosind procedee diferite sau printr-un același procedeu, dar utilizînd agregate cu performanțe diferite.

Problema care se pune este aceea de a ști care dintre utilaje asigură o stabilitate mai mare a caracteristicilor produsului finit, adică în ultimă instanță, care este mai precis.

În mod analog se pune problema și în cazul a două procedee tehnice diferite folosite pe un același agregat.

Scanned with OKEN Scanner

Să mai observăm că la folosirea unui agregat se constată anumite variații ale preciziei sale în timp și atunci se poate pune problema comparării preciziei la două momente diferite.

Fie X și Y două variabile aleatoare independente, caracteristici de calitate, ale căror densități $f(x; \theta_1)$, $f(y; \theta_2)$ cu aceeași formă cunoscută însă parametrii θ_1 și θ_2 sînt necunoscuți.

Acești parametri necunoscuți pot fi interpretați ca parametri ai preciziei din repartițiile celor două caracteristici de calitate determinate de desfășurarea independentă a două procese tehnologice ce au ca rezultat un același produs.

Trebuie să verificăm ipoteza compusă $H_0 : \theta_1 < \theta_2$ față de alternativa $H_1 : \theta_1 \geqslant \theta_2$.

Ca exemplu, X poate reprezenta rezistența la încovoiere a unei bare metalice, iar $\theta_1 = \sigma_1$ mărimea abaterii standard prin folosirea primului procedeu de fabricație a barelor, iar $\theta_2 = \sigma_2$ este mărimea abaterii standard pentru caracteristica de calitate Y ce reprezintă rezistența de încovoiere a barei, obținută însă cu cel de-al doilea procedeu tehnologic.

Este natural să se prefere procedeul care realizează abaterea standard cea mai mică, deci problema optării se reduce la verificarea ipotezei statistice $H_0 : \sigma_1 < \sigma_2$ față de alternativa $H_1 : \sigma_1 \geqslant \sigma_2$, alegîndu-se prima metodă de fabricație dacă se verifică H_0 sau cea de-a doua dacă se verifică H_1.

Problema se poate pune în mod analog și pentru caracteristici calitative discrete.

De exemplu, dîndu-se două loturi de produse de același fel, obținute însă pe două utilaje diferite, se pune problema stabilirii lotului de calitatea cea mai bună.

Dacă „p“ reprezintă probabilitatea ca un obiect extras la întîmplare din primul lot să fie defect iar π în mod analog pentru cel de-al doilea lot, problema se reduce la verificarea ipotezei statistice $H_0 : p < \pi$ față de alternativa $H_1 : p \geqslant \pi$.

Cei doi parametri p și π sînt necunoscuți, dar se știe că ei sînt caracterizați de repartiția binomială.

În cazul nostru trebuie înlocuită verificarea ipotezei compuse $H_0 : \theta_1 < \theta_2$ față de alternativa $H_1 : \theta_1 \geqslant \theta_2$ cu verificarea ipotezei simple

$H_0' : \theta_1 = \theta_1^0$, $\theta_2 = \theta_2^0$ față de alternativa $H_1' : \theta_1 = \theta_2^0$, $\theta_2 = \theta_1^0$, unde θ_1^0 și θ_2^0 se determină din condiția $v(\theta_1^0; \theta_2^0) = -\delta(\theta_1^0 < \theta_2^0)$.

Scanned with OKEN Scanner

Testul secvențial pentru verificarea ipotezei H_0 față de alternativa H_1 ia decizia în conformitate cu procedeul descris în § 2.

O mărime deosebit de utilă pentru informațiile pe care le furnizează, în cazul verificării, este funcția caracteristică operatoare $L(\theta_1; \theta_2)$, care se definește ca probabilitatea admiterii ipotezei H_0 cînd adevărata valoare a parametrilor este θ_1 și respectiv θ_2. Ea se exprimă prin formula

$$L(\theta_1; \theta_2) = \frac{A^{h(\theta_1, \theta_2)} - 1}{A^{h(\theta_1, \theta_2)} - B^{h(\theta_1, \theta_2)}}$$

unde $A = \frac{1-\beta}{\alpha}$, $B = \frac{\beta}{1-\alpha}$ iar $h(\theta_1, \theta_2)$ se definește astfel ca valoarea medie a expresiei $[f(x; \theta_2)/f(x; \theta_1)]^{h(\theta_1, \theta_2)}$ să fie egală cu 1.

În cazul repartițiilor de tipul [1)]

$$f(x; \theta) = \exp[A(x)\, r(\theta) + B(x) + S(\theta)]$$

funcția $h(\theta_1, \theta_2)$ se determină cu ajutorul relației

$$h(\theta_1, \theta_2) = -\frac{1}{\delta} v(\theta_1, \theta_2).$$

În cazul în care variabilele X, Y sînt repartizate normal $N(0, \sigma_1^2)$ și $N(0, \sigma_2^2)$ avem

$$h(\sigma_1, \sigma_2) = \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) \Big/ \left(\frac{1}{(\sigma_2^0)^2} - \frac{1}{(\sigma_1^0)^2}\right).$$

Dacă variabilele au repartiții binomiale, caracterizate de parametrii p și π, atunci se obține

$$h(p, \pi) = \ln \frac{\pi(1-p)}{p(1-\pi)} \Big/ \ln \frac{\pi_0(1-p_0)}{p_0(1-\pi_0)}.$$

Cantitatea $\delta > 0$ ne permite determinarea valorilor θ_1^0 și θ_2^0 $(\theta_1^0 < \theta_2^0)$ cu ajutorul cărora se construiește testul secvențial.

Metoda de verificare trebuie să satisfacă următoarele condiții:

1) probabilitatea respingerii ipotezei H_0' nu trebuie să fie mai mare decît riscul de genul I(α) în tot domeniul definit de inegalitatea

$$v(\theta_1, \theta_2) \leqslant \delta;$$

[1)] pentru detalii a se vedea [31]

Scanned with OKEN Scanner

2) probabilitatea acceptării ipotezei H_0' nu trebuie să depășească riscul de genul II(β) în tot domeniul definit de inegalitatea:

$$v(\theta_1, \theta_2) \leqslant \delta.$$

Aceste condiții odată îndeplinite, se aleg valorile θ_1^0 și θ_2^0 în așa fel încît

$$v(\theta_1^0, \theta_2^0) = -\delta.$$

Se observă că pentru un $\delta > 0$ dat, există o infinitate de valori θ_1^0, θ_2^0 în așa fel încît $v(\theta_1^0, \theta_2^0) = -\delta$.

Pentru a se înlătura această nedeterminare, se procedează în felul următor: se dă de la bun început mărimea η a diferenței admise dintre cei doi parametri θ_1 și θ_2, mărime ce rezultă din cerințele practice.

Prin urmare, nu se poate lua nici o decizie asupra ipotezei $\theta_1 < \theta_2$ dacă avem $|\theta_1 - \theta_2| < \eta$ dar putem decide în cazul în care $|\theta_1 - \theta_2| \geqslant \eta$.

Cu aceasta se găsește că $\theta_2^0 = \theta_1^0 + \eta$, iar θ_1^0 se determină în mod unic din condiția $v(\theta_1, \theta_1 + \eta) = \text{maxim}$.

Calculele arată că maximul funcției

$$v(\sigma_1^2, \sigma_2^2) = \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)$$

este atins pentru

$(\sigma_1^0)^2 = \frac{\eta}{2}$, $(\sigma_2^0)^2 = (\sigma_1^0)^2 + \eta$ iar maximul funcției $v(p, \pi) =$

$= \ln \frac{p(1-\pi)}{\pi(1-p)}$ este atins pentru $p_0 = \frac{1}{2}(1 - \eta)$ cu $\pi_0 = p_0 + \eta$.

Aplicații ale testului secvențial la controlul statistic

a) *Cazul repartiției binomiale a defectelor (volumul lotului $N \geqslant 5\,000$ unități).*

Fie p procentul de rebuturi din primul lot iar π procentul de rebuturi din cel de-al doilea lot.

A arăta care dintre cele două utilaje (sau secții) este mai precis revine la a verifica ipoteza $H_0 : p < \pi$ față de alternativa $H_1 : p \geqslant \pi$ ceea ce revine la a verifica ipoteza $H_0' : p = p_0$, $\pi = \pi_0$ față de alternativa $H_1' : p = \pi_0$, $\pi = p_0$.

Scanned with OKEN Scanner

Verificarea se face pas cu pas calculînd pentru fiecare pereche (x_i, y_i) de observaţii, raportul r_n. Dacă se notează cu d_n şi ν_n respectiv numărul de piese defecte din selecţiile $(x_1, \ldots, x_n)$ şi $(y_1, \ldots, y_n)$ atunci se obţine

$$r_n = \exp[\delta(d_n - \nu_n)].$$

Aşadar:

— se admite ipoteza H_0 dacă $d_n - \nu_n \leqslant \frac{1}{\delta} \ln \frac{\beta}{1-\alpha}$,

— se respinge ipoteza H_0 dacă $d_n - \nu_n \leqslant \frac{1}{\delta} \ln \frac{1-\beta}{\alpha}$;

— se continuă experimentul dacă:

$$\frac{1}{\delta} \ln \frac{\beta}{1-\alpha} < d_n - \nu_n < \frac{1}{\delta} \ln \frac{1-\beta}{\alpha}.$$

Din considerente practice s-a ales diferenţa admisă dintre p şi π

$$|p - \pi| \geqslant 0{,}01, \quad \text{deci} \quad \eta = 0{,}01.$$

Se găseşte deci $p_0 = 0{,}495$, $\pi_0 = 0{,}505$, $\delta = 0{,}04$.

Dacă se va alege $\alpha = \beta = 0{,}10$, atunci vom avea

$$\ln \frac{\beta}{1-\alpha} = -2{,}197, \quad \ln \frac{1-\beta}{\alpha} = 2{,}197,$$

$$\frac{1}{\delta} \ln \frac{\beta}{1-\alpha} = -55, \quad \frac{1}{\delta} \ln \frac{1-\beta}{\alpha} = 55.$$

Aşadar dacă $d_n - \nu_n \leqslant -55$ se va accepta H_0.

Dacă $d_n - \nu_n \geqslant 55$ atunci se va accepta H_1 (adică lotul al doilea este mai bun).

Se va continua experimental dacă $-55 < d_n - \nu_n < 55$.

b) *Cazul repartiţiei normale a caracteristicii măsurate.*

În cazul în care caracteristicile de calitate X şi Y sînt repartizate normal $N_1(0, \sigma_1^2)$ şi $N_2(0, \sigma_2^2)$ metoda se aplică, în mod analog.

Să presupunem de exemplu că X şi Y reprezintă diametrele interioare a două piese de acelaşi tip fabricate însă de maşini diferite şi că pentru comparare avem la dispoziţie două loturi

cu volum suficient de mare. Parametrii de precizie vor fi cele două dispersii σ_1^2 și σ_2^2. Agregatul de la care s-a obținut lotul cu dispersia cea mai mică va fi cel mai precis.

Deci în acest caz trebuie verificată $H_0 : \sigma_1^2 < \sigma_2^2$ față de alternativa $H_1 : \sigma_1^2 \geqslant \sigma_2^2$ sau echivalenta ei $H_0' : \sigma_1^2 = (\sigma_1^0)^2; \sigma_2^2 = (\sigma_2^0)^2$ față de alternativa $H_1' : \sigma_1^2 = (\sigma_2^0)^2$, $\sigma_2^2 = (\sigma_1^0)^2$.

Ținînd cont de forma densităților de probabilitate se obține

$$r_n = \exp\left[\delta \sum_1^n (x_i^2 - y_i^2)\right].$$

Așadar:

1) se acceptă H_0 dacă $\sum_1^n (x_i^2 - y_i^2) \leqslant \frac{1}{\delta} \ln \frac{\beta}{1-\alpha}$;

2) se respinge H_0 dacă $\sum_1^n (x_i^2 - y_i^2) \geqslant \frac{1}{\delta} \ln \frac{1-\beta}{\alpha}$,

3) se continuă experimentul, dacă:

$$\frac{1}{\delta} \ln \frac{\beta}{1-\alpha} < \sum_1^n (x_i^2 - y_i^2) < \frac{1}{\delta} \ln \frac{1-\beta}{\alpha}.$$

În aceste relații x_i și y_i reprezintă cotele exprimate în microni ale diametrelor pieselor din proba „i“ extrasă din cele două loturi, iar

$$\delta = \frac{1}{2}\left[\frac{1}{(\sigma_1^0)^2} - \frac{1}{(\sigma_2^0)^2}\right].$$

Metodologia prezentată mai sus, poate fi folosită la efectuarea controlului statistic de recepție în cazul în care sosește la control un număr mare de loturi de produse identice.

Pentru aceasta, controlul decurge în felul următor: se alege un lot pentru care se cunoaște bine prin verificări anterioare valoarea exactă a procentului de rebut „p“. Lotul se alege astfel ca „p“ să aibă o valoare limită maximă admisibilă asupra căruia cad de acord atît furnizorul cît și beneficiarul.

Un astfel de lot poartă numele de „lot martor“. Celelalte loturi se compară cu lotul martor după cele expuse anterior.

Scanned with OKEN Scanner

Se vor accepta loturile pentru care $p_i \leqslant p$, iar loturile pentru care $p_i > p$ vor fi respinse supunîndu-se controlului total pentru trierea exemplarelor defecte.

Exemplul 1. (D. V. Iliescu, Gh. Obreja și V. Gh. Vodă: *Metodă de verificare a ipotezelor statistice în controlul proceselor tehnologice în industrie*, „Revista de statistică" nr. 4/1971 pp. 44—54).

Se compară performanțele a două sisteme de filare a poliamidelor. Filarea fibrelor din granule poliamidice se realizează prin două metode: filare pe capete de filare individuale încălzite electric și filare prin sisteme comune (grinzi de filare).

Unul dintre parametrii importanți care ar contribui la luarea deciziei în folosirea unuia dintre cele două procedee este uniformitatea caracteristicilor fizico-mecanice ale produsului finit — în cazul de față fibrele poliamidice.

Se urmărește să se vadă care dintre cele două tehnologii asigură o împrăștiere mai mică a valorilor la caracteristica „alungire" a fibrelor poliamidice.

Fie: X — caracteristica de calitate aferentă liniei de filare ce folosește sistemul de filare cu grinzi de filare și Y — caracteristica de calitate aferentă liniei de filare ce folosește sistemul de filare cu capete de filare individuale.

Ambele caracteristici se presupun repartizate normal. Parametrii necunoscuți sînt dispersiile σ_1^2 și σ_2^2.

Deoarece în cazul de față valoarea lui η este greu de stabilit în mod direct, fiindcă reprezintă o diferență maximă între două dispersii sub care s-ar considera practic că cele două linii tehnologice sînt la fel de precise, ea se estimează făcînd uz de diferența maximă admisibilă Φ apreciată din practică între abaterile standard și mărimea:

$$\lambda = \frac{1}{6}(T_s - T_i)$$

care dau pentru cele două linii tehnologice o valoare estimativă asupra ordinului de mărime al lui σ_1 și σ_2.

Scanned with OKEN Scanner

Atunci, η se determină din relaţia $\eta = 2\lambda\Phi + \Phi^2$.

Tabelul 10.1 sintetizează datele obţinute pentru caracteristica „alungire".

Tabelul 10.1

Nr.	u_i	$\bar{u}_i$	w_i	$\bar{w}_i$	$x_i = u_i - \bar{u}_i$	$y_i = w_i - w_i$	x_i^2	y_i^2	$x_i^2 - y_i^2$	$\Sigma(x_i^2 - y_i^2)$
1	36,8	36,8	38,4	38,4	0	0	0	0	0	0
2	41,2	39,0	37,8	38,1	2,2	—0,3	4,84	0,09	4,75	4,75
3	37,5	38,5	39,4	38,5	—1,0	0,9	1,00	1,81	0,19	4,94
4	34,7	37,5	46,5	40,5	—2,8	6,0	7,84	36,00	—28,16	—23,02
5	34,7	36,9	32,6	38,8	—2,2	—6,3	4,84	39,69	—34,85	—57,87

Calculele au furnizat

$\lambda = 3{,}5$, $\Phi = 1$, $\eta = 8$, $(\sigma_1^0)^2 = 4$, $(\sigma_2^0)^2 = 12$, $\delta = \frac{1}{12}$ iar pentru $\alpha = 0{,}05$ şi $\beta = 0{,}10$

$$\frac{1}{\delta}\ln\frac{\beta}{1-\alpha} = -27, \quad \frac{1}{\delta}\ln\frac{1-\beta}{\alpha} = 34{,}9.$$

Deoarece

$$\sum_1^5 (x_i^2 - y_i^2) < \frac{1}{\delta}\ln\frac{\beta}{1-\alpha}$$

se acceptă ipoteza $H_0 : \sigma_1^2 < \sigma_2^2$.

Acest fapt petrecîndu-se după un număr mic de observaţii rezultă că tehnologia (X) asigură o uniformitate mult mai mare a caracteristicii „alungire" faţă de tehnologia (Y).

Exemplul 2. (V. Gh. Vodă: *Calculul numărului mediu de probe necesar în cazul unor teste secvenţiale aplicate în chimie*, „Materiale plastice" 1972).

Scanned with OKEN Scanner

În această lucrare se calculează numărul mediu de probe necesar verificării ipotezei $H_0 : \sigma_1 < \sigma_2$ față de alternativa $H_1 : \sigma_1 \geqslant \sigma_2$ chiar în cazul citat la exemplul 1. Pentru calculul numărului mediu de probe (care în cazul testelor secvențiale este o variabilă aleatoare) se folosește formula lui Wald modificată și anume

$$E_{\sigma_1;\sigma_2}(n) = \frac{L(\sigma_1, \sigma_2) \ln B + (1 - L(\sigma_1;\sigma_2)) \ln A}{\delta(\sigma_1^2 - \sigma_2^2)}$$

unde $L(\sigma_1; \sigma_2)$ este funcția caracteristică operatoare a testului a cărei formulă s-a dat în acest paragraf.

Ținînd cont de elementele de la exemplul 1 se calculează $A = 18$, $B \simeq 0{,}1$, $\ln A \simeq 3$, $\ln B = -2$, $\sigma(\sigma_1^2 - \sigma_2^2) \simeq 0{,}6$ $L(\sigma_1; \sigma_2) \simeq 0{,}95$.

Înlocuind în formula lui Wald se obține

$$E_{\sigma_1;\sigma_2}(n) \simeq 5.$$

Acest rezultat arată că în medie, după cinci măsurători experimentul poate fi oprit, putîndu-se lua decizia cu privire la ipoteza H_0.

Acest lucru este confirmat de înseși calculele de la exemplul 1.

Exemplul 3. (Al. C. Dorin și V. Gh. Vodă: *Compararea duratei medii de funcționare a sculelor așchietoare ale unui centru de prelucrare prin metode statistice*, „Revista de Statistică" nr. 4/1972).

Durata de funcționare a sculelor așchietoare ale unui centru de prelucrare este descrisă de repartiția exponențială

$$f(x;\lambda) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right), \quad x > 0, \quad \lambda > 0.$$

Se știe că $M(X) = \lambda$, iar $D^2(X) = \lambda^2$. Așadar parametrul λ semnifică durata medie de funcționare a sculelor respective.

Elementele necesare verificării ipotezei $H_0 : \theta = M(X) < \omega = M(Y)$ față de alternativa $H_1 : \theta = M(X) \geqslant \omega = M(Y)$, care

exprimă tocmai o comparație între duratele medii de viață ale celor două tipuri de scule, sînt următoarele:

$$h(\theta,\ \omega) = -\ \frac{\theta_0\omega_0(\omega - \theta)}{\theta\omega(\omega_0 - \theta_0)},$$

$$L(\theta\,;\,\omega) = [A^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}} - 1] \,/\, [A^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}} - B^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0}}],$$

$$M_{\theta;\,\omega}(n) = \frac{\theta_0\omega_0}{(\omega_0 - \theta_0)\ (\theta - \omega)\left[A^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}} - B^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}}\right]} \times$$

$$\times\left[A^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}} \ln B - B^{\frac{\theta_0\omega_0(\theta-\omega)}{\theta\omega(\omega_0-\theta_0)}} \ln A + \ln\frac{A}{B}\right].$$

Scanned with OKEN Scanner

BIBLIOGRAFIE

1. Brownlee, K. A., *Statistical Theory and Metodology in Science and Engineering*, John Wiley, New-York, London, 1960.
2. Chernoff, H., *Lincoln, M., Elementary Decision Theory*, John Wiley, New York, 1959.
3. Ciucu, G., Craiu V., *Teoria estimației și verificarea ipotezelor statistice*, Editura didactică și pedagogică, București 1969.
4. Ciucu, G., Craiu, V., *Introducere în teoria probabilităților și statistică matematică*, Editura didactică și pedagogică, București, 1971.
5. Cochran, W., G., *The χ^2 test of goodness of fit*, Ann. Math. Stat. 23 (1952).
6. Cornifield, Jerome, John. W. Tukey, *Average values of mean squares in factorials*. Ann. Math. Stat 27 (1956).
7. Cramer, H., *Mathematical Methods of statistics*, Princeton University Press, 1951.
8. Dixon, W. J., Massey, F., Jr., *Introduction to statistical Analysis*, Ed. 2, Mc. Graw Book Company, New York, 1957.
9. Feller, W. *On introduction to Probability Theory and its applications*, vol. 1, John Wiley, New York, 1957.
10. František Zítek. *Nekteré analogie necentralniho t-testu*, „Aplikace Matematiky", 2 (1957), pp. 38—46.
11. Fraser D. A., *Nonparametric Methods in Statistics*, John Wiley, New York, 1957.
12. Gnedenko, B., *The Theory of Probability*, Mir. Publishers, Moscou, 1969.
13. Gottfried e Noether, *Elements of nonparametric statistics*, John Wiley, New York.
14. Guttman, I., Wilks, S. S., *Introductory Engineering Statistics* John Wiley, New York, 1965.
15. Hays, W. L., *Statistics*, London, 1969.
16. Hellwig, Z. *Test zgodnosci dla malej prolki*, „Przeglad Statysticzny" 2 (1965).

Scanned with OKEN Scanner

37. Scheffé, H. *Fitting straight lines when one variable is controlled*, Journal of the American Statistical Association, 53 (1958).
38. Snedecor, G. W., *Metode statistice aplicate în cercetările de agricultură şi biologie* (trad. din lb. engleză), Editura didactică şi pedagogică, Bucureşti, 1968.
39. Tucker, H. G., *On introduction to probability and mathematical statistics*, Academic Press Textbooks in Mathematics, New York, 1962.
41. Zbigniew Kowalsky, *O testach do weryfikacji hipotezy o normalnym rozkladzie w przypadku prólki o maley liczności* „Przeglad Statisticzny", tom 17, 2/1970) pp. 221—219.
42. Van der Waerden, *Mathematische Statistik*, Springer-Verlag, Berlin, 1957.
43. Wald, A. *Statistical Decision Functions*, John Wiley & Sons., New York, 1950.
44. Yule, G. U., Kendall, M. G. *Introducere în teoria statisticii* (trad. din lb. engleză) Ediţia a XIV-a, Editura ştiinţifică, Bucureşti, 1969.

Scanned with OKEN Scanner

Valorile funcției $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$ *Tabelul 1.*

t	0	1	2	3	4	5	6	7	8	9
—0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
—0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
—0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
—0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
—0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
—0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
—0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
—0,7	2420	2389	2358	2327	2297	2266	2236	2206	2177	2148
—0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
—0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
—1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
—1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
—1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
—1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
—1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
—1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
—1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
—1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
—1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
—1,90	0288	0281	0274	0268	0262	0256	0250	0244	0239	0233
—2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
—2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
—2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
—2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
—2,4	0082	0080	0078	0075	0073	0071	0069	0068	0058	0064
—2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
—2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
—2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
—2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
—2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014

$t =$	—3,0	—3,1	—3,2	—3,3	—3,4	—3,5	—3,6	—3,7	—3,8	—3,9
$\Phi(t) =$	0,0013	0,0010	0,0007	0,0005	0,0003	0,0002	0,0002	0,0001	0,0001	0,0000

Scanned with OKEN Scanner

0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6338	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7965	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8331	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9687	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986

t =	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$=	0,9987	0,9990	0,9993	0,9995	0,9997	0,9998	0,9998	0,9999	0,9999	1,0000

Tabelele 1 2, 3a 5a, 5b şi 6 sînt luate din [39]
Tabela 4 este luată din]12]
Tabelele 5c şi 7 sînt luate din Henry L. Alder, Edward B. Roessler, *Introduction to Probability and Statistics*, Four Edition, Warner House, Folkestone, Kent, 1968.

Scanned with OKEN Scanner

Tabelul 2

Valorile funcției inverse $\psi(x)$ a lui $\Phi(t)$

x	0	1	2	3	4	5	6	7	8	9
0,00	—∞	—3,09	—2,88	—2,75	—2,65	—2,58	—2,51	—2,46	—2,41	—2,37
0,01	—2,33	—2,29	—2,26	—2,23	—2,20	—2,17	—2,14	—2,12	—2,10	—2,07
0,02	—2,05	—2,03	—2,01	—2,00	—1,98	—1,96	—1,94	—1,93	—1,91	—1,90
0,03	—1,88	—1,87	—1,85	—1,84	—1,83	—1,81	—1,80	—1,79	—1,77	—1,76
0,04	—1,75	—1,74	—1,73	—1,72	—1,71	—1,70	—1,68	—1,67	—1,66	—1,65
0,05	—1,64	—1,64	—1,63	—1,62	—1,61	—1,60	—1,59	—1,58	—1,57	—1,56
0,06	—1,55	—1,55	—1,54	—1,53	—1,52	—1,51	—1,51	—1,50	—1,49	—1,48
0,07	—1,48	—1,47	—1,46	—1,45	—1,45	—1,44	—1,43	—1,43	—1,42	—1,41
0,08	—1,41	—1,40	—1,39	—1,39	—1,38	—1,37	—1,37	—1,36	—1,35	—1,35
0,09	—1,34	—1,33	—1,33	—1,32	—1,32	—1,31	—1,30	—1,30	—1,29	—1,29
0,10	—1,28	—1,28	—1,27	—1,26	—1,26	—1,25	—1,25	—1,24	—1,24	—1,23
0,11	—1,23	—1,22	—1,22	—1,21	—1,21	—1,20	—1,20	—1,19	—1,19	—1,18
0,12	—1,18	—1,17	—1,17	—1,16	—1,16	—1,15	—1,15	—1,14	—1,14	—1,13
0,13	—1,13	—1,12	—1,12	—1,11	—1,11	—1,10	—1,10	—1,19	—1,09	—1,09
0,14	—1,08	—1,08	—1,07	—1,07	—1,06	—1,06	—1,05	—1,05	—1,05	—1,04
0,15	—1,04	—1,03	—1,03	—1,02	—1,02	—1,01	—1,01	—1,01	—1,00	—1,00
0,16	—0,99	—0,99	—0,99	—0,98	—0,98	—0,97	—0,97	—0,97	—0,96	—0,96
0,17	—0,95	—0,95	—0,95	—0,94	—0,94	—0,93	—0,93	—0,93	—0,92	—0,92
0,18	—0,92	—0,91	—0,91	—0,90	—0,90	—0,90	—0,89	—0,89	—0,89	—0,88
0,19	—0,88	—0,87	—0,87	—0,87	—0,86	—0,86	—0,86	—0,85	—0,85	—0,85
0,20	—0,84	—0,84	—0,83	—0,83	—0,83	—0,82	—0,82	—0,82	—0,81	—0,81
0,21	—0,81	—0,80	—0,80	—0,80	—0,79	—0,79	—0,79	—0,78	—0,78	—0,78

Scanned with OKEN Scanner

0,22	−0,77	−0,77	−0,77	−0,76	−0,76	−0,76	−0,75	−0,75	−0,75	−0,74
0,23	−0,74	−0,74	−0,73	−0,73	−0,73	−0,72	−0,72	−0,72	−0,71	−0,71
0,24	−0,71	−0,70	−0,70	−0,70	−0,69	−0,69	−0,69	−0,68	−0,68	−0,68
0,25	−0,67	−0,67	−0,67	−0,67	−0,66	−0,66	−0,66	−0,65	−0,65	−0,65
0,26	−0,64	−0,64	−0,64	−0,63	−0,63	−0,63	−0,63	−0,62	−0,62	−0,62
0,27	−0,61	−0,61	−0,61	−0,60	−0,60	−0,60	−0,59	−0,59	−0,59	−0,59
0,28	−0,58	−0,58	−0,58	−0,57	−0,57	−0,57	−0,57	−0,56	−0,56	−0,56
0,29	−0,55	−0,55	−0,55	−0,54	−0,54	−0,54	−0,54	−0,53	−0,53	−0,53
0,30	−0,52	−0,52	−0,52	−0,52	−0,51	−0,51	−0,51	−0,50	−0,50	−0,50
0,31	−0,50	−0,49	−0,49	−0,49	−0,48	−0,48	−0,48	−0,48	−0,47	−0,47
0,32	−0,47	−0,46	−0,46	−0,46	−0,46	−0,45	−0,45	−0,45	−0,45	−0,44
0,33	−0,44	−0,44	−0,43	−0,43	−0,43	−0,43	−0,42	−0,42	−0,42	−0,42
0,34	−0,41	−0,41	−0,41	−0,40	−0,40	−0,40	−0,40	−0,39	−0,39	−0,39
0,35	−0,39	−0,38	−0,38	−0,38	−0,37	−0,37	−0,37	−0,37	−0,36	−0,36
0,36	−0,36	−0,36	−0,35	−0,35	−0,35	−0,35	−0,34	−0,34	−0,34	−0,33
0,37	−0,33	−0,33	−0,33	−0,32	−0,32	−0,32	−0,32	−0,31	−0,31	−0,31
0,38	−0,31	−0,30	−0,30	−0,30	−0,30	−0,29	−0,29	−0,29	−0,28	−0,28
0,39	−0,28	−0,28	−0,27	−0,27	−0,27	−0,27	−0,26	−0,26	−0,26	−0,26
0,40	−0,25	−0,25	−0,25	−0,25	−0,24	−0,24	−0,24	−0,24	−0,23	−0,23
0,41	−0,23	−0,23	−0,22	−0,22	−0,22	−0,21	−0,21	−0,21	−0,21	−0,20
0,42	−0,20	−0,20	−0,20	−0,19	−0,19	−0,19	−0,19	−0,18	−0,18	−0,18
0,43	−0,18	−0,17	−0,17	−0,17	−0,17	−0,17	−0,16	−0,16	−0,16	−0,15
0,44	−0,15	−0,15	−0,15	−0,14	−0,14	−0,14	−0,14	−0,13	−0,13	−0,13
0,45	−0,13	−0,12	−0,12	−0,12	−0,12	−0,11	−0,11	−0,11	−0,11	−0,10
0,46	−0,10	−0,10	−0,10	−0,09	−0,09	−0,09	−0,09	−0,08	−0,08	−0,08
0,47	−0,08	−0,07	−0,07	−0,07	−0,07	−0,06	−0,06	−0,06	−0,06	−0,05

Scanned with OKEN Scanner

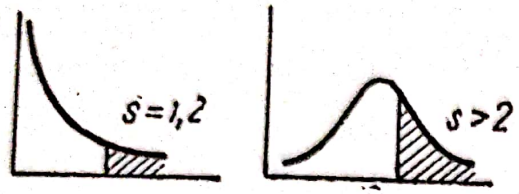
Tabelul 2 (continuare)

x	0	1	2	3	4	5	6	7	8	9
0,48	—0,05	—0,05	—0,05	—0,04	—0,04	—0,04	—0,04	—0,03	—0,03	—0,03
0,49	—0,03	—0,02	—0,02	—0,02	—0,02	—0,01	—0,01	—0,01	—0,01	—0,00
0,50	—0,00	—0,00	—0,01	—0,01	—0,01	—0,01	—0,02	—0,02	—0,02	—0,02
0,51	—0,03	—0,03	—0,03	—0,03	—0,04	—0,04	—0,04	—0,04	—0,05	—0,05
0,52	—0,05	—0,05	—0,06	—0,06	—0,06	—0,06	—0,07	—0,07	—0,07	—0,07
0,53	—0,08	—0,08	—0,08	—0,08	—0,09	—0,09	—0,09	—0,09	—0,10	—0,10
0,54	—0,10	—0,10	—0,11	—0,11	—0,11	—0,11	—0,12	—0,12	—0,12	—0,12
0,55	—0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
0,56	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
0,57	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
0,58	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
0,59	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
0,60	0,25	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,27	0,28
0,61	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
0,62	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
0,63	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
0,64	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
0,65	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
0,66	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,41
0,67	0,44	0,44	0,45	0,45	0,45	0,45	0,46	0,46	0,46	0,46
0,68	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
0,69	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52
0,70	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
0,71	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
0,72	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
0,73	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64

Scanned with OKEN Scanner

0,74	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
0,75	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
0,76	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
0,77	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
0,78	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
0,79	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
0,80	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
0,81	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
0,82	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95
0,83	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
0,84	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
0,85	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
0,86	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
0,87	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
0,88	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
0,89	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,28
0,90	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
0,91	1,34	1,35	1,35	1,36	1,37	1,37	1,38	1,39	1,39	1,40
0,92	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
0,93	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
0,94	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
0,95	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
0,96	1,75	1,76	1,77	1,79	1,80	1,81	1,83	1,84	1,85	1,87
0,97	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
0,98	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
0,99	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Scanned with OKEN Scanner



Tabelul 3

Valorile lui χ^2 (Testul χ^2)

α reprezintă aria haşurată pentru valorile lui χ^2 date mai jos

s	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}001$	s	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
1	3,84	6,63	10,8	51	68,7	77,4	88,0
2	5,99	9,21	13,8	52	69,8	78,6	89,3
3	7,81	11,3	16,3	53	71,0	79,8	90,6
4	9,49	13,3	18,5	54	72,2	81,1	91,9
5	11,1	15,1	20,5	55	73,3	82,3	93,2
6	12,6	16,8	22,5	56	74,5	83,5	94,5
7	14,1	18,5	24,3	57	75,6	84,7	95,8
8	15,5	20,1	26,1	58	76,8	86,0	97,0
9	16,9	21,7	27,9	59	77,9	87,2	98,3
10	18,3	23,2	29,6	60	79,1	88,4	99,6
11	19,7	24,7	31,3	61	80,2	89,6	100,9
12	21,0	26,2	32,9	62	81,4	90,8	102,2
13	22,4	27,7	34,5	63	82,5	92,0	103,4
14	23,7	29,1	36,1	64	83,7	93,2	104,7
15	25,0	30,6	37,7	65	84,8	94,4	106,0
16	26,3	32,0	39,3	66	86,0	95,6	107,3
17	27,6	33,4	40,8	67	87,1	96,8	108,5
18	28,9	34,8	42,3	68	88,3	98,0	109,8
19	30,1	36,2	43,8	69	89,4	99,2	111,1
20	31,4	37,6	45,3	70	90,5	100,4	112,3
21	32,7	38,9	46,8	71	91,7	101,6	113,6
22	33,9	40,3	48,3	72	92,8	102,8	114,8
23	35,2	41,6	49,7	73	93,9	104,0	116,1

Scanned with OKEN Scanner

24	36,4	43,0	51,2	74	95,1	105,2	117,3
25	37,7	44,3	52,6	75	96,2	106,4	118,6
26	38,9	45,6	54,1	76	97,4	107,6	119,9
27	40,1	47,0	55,5	77	98,5	108,8	121,1
28	41,3	48,3	56,9	78	99,6	110,0	122,3
29	42,6	59,6	58,3	79	100,7	111,1	123,6
30	43,8	50,9	59,7	80	101,9	112,3	124,8
31	45,0	52,2	61,1	81	103,0	113,5	126,1
32	46,2	53,5	62,5	82	104,1	114,7	127,3
33	47,4	54,8	63,9	83	105,3	115,9	128,6
34	48,6	56,1	65,2	84	106,4	117,1	129,8
35	49,8	57,3	66,6	85	107,5	118,2	131,0
36	51,0	58,6	68,0	86	108,6	119,4	123,3
37	52,2	59,9	69,3	87	109,8	120,6	133,5
38	53,4	61,2	70,7	88	110,9	121,8	134,7
39	54,6	62,4	72,1	89	112,0	122,9	136,0
40	55,8	63,7	73,4	90	113,1	124,1	137,2
41	56,9	65,0	74,7	91	114,3	125,3	138,4
42	58,1	66,2	76,1	92	115,4	126,5	139,7
43	59,3	67,5	77,4	93	116,5	127,6	140,9
44	60,5	68,7	78,7	94	117,6	128,8	142,1
45	61,7	70,0	80,1	95	118,8	130,0	143,3
46	62,8	71,2	81,4	96	119,9	131,1	144,6
47	64,0	72,4	82,7	97	121,0	132,3	145,8
48	65,2	73,7	84,0	98	122,1	133,5	147,0
49	66,3	74,9	85,4	99	123,2	134,6	148,2
50	67,5	76,2	86,7	100	124,3	135,8	149,4

Scanned with OKEN Scanner

Tabelul 4

Valorile funcției $K(\lambda) = \sum_{-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$

(testul asimptotic al lui Kolmogorov)

λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$
0,34	0,0001	0,71	0,3054	1,08	0,8061	1,45	0,9701
0,35	0,0003	0,72	0,3222	1,09	0,8143	1,46	0,9718
0,36	0,0005	0,73	0,3391	1,10	0,8222	1,47	0,9734
0,37	0,0008	0,74	0,3559	1,11	0,8299	1,48	0,9749
0,38	0,0012	0,75	0,3728	1,12	0,8373	1,49	0,9764
0,39	0,0019	0,76	0,3896	1,13	0,8445	1,50	0,9777
0,40	0,0028	0,77	0,4063	1,14	0,8513	1,51	0,9790
0,41	0,0039	0,78	0,4230	1,15	0,8580	1,52	0,9803
0,42	0,0054	0,79	0,4395	1,16	0,8644	1,53	0,9814
0,43	0,0073	0,80	0,4558	1,17	0,8706	1,54	0,9825
0,44	0,0097	0,81	0,4720	1,18	0,8765	1,55	0,9836
0,45	0,0125	0,82	0,4880	1,19	0,8822	1,56	0,9846
0,46	0,0160	0,83	0,5038	1,20	0,8877	1,57	0,9855
0,47	0,0200	0,84	0,5193	1,21	0,8930	1,58	0,9864
0,48	0,0246	0,85	0,5346	1,22	0,8981	1,59	0,9872

Scanned with OKEN Scanner

0,49	0,0300	0,86	0,5497	1,23	0,9029	1,60	0,9880
0,50	0,0360	0,87	0,5645	1,24	0,9076	1,61	0,9887
0,51	0,0428	0,88	0,5790	1,25	0,9121	1,62	0,9894
0,52	0,0503	0,89	0,5933	1,26	0,9164	1,63	0,9901
0,53	0,0585	0,90	0,6072	1,27	0,9205	1,64	0,9907
0,54	0,0674	0,91	0,6209	1,28	0,9245	1,65	0,9913
0,55	0,0771	0,92	0,6342	1,29	0,9282	1,66	0,9919
0,56	0,0875	0,93	0,6473	1,30	0,9319	1,67	0,9924
0,57	0,0986	0,94	0,6600	1,31	0,9353	1,68	0,9929
0,58	0,1103	0,95	0,6725	1,32	0,9386	1,69	0,9933
0,59	0,1227	0,96	0,6846	1,33	0,9418	1,70	0,9938
0,60	0,1357	0,97	0,6964	1,34	0,9448	1,71	9,9942
0,61	0,1492	0,98	0,7079	1,35	0,9477	1,72	0,9946
0,62	0,1632	0,99	0,7191	1,36	0,9505	1,73	0,9949
0,63	0,1777	1,00	0,7300	1,37	0,9531	1,74	0,9953
0,64	0,1926	1,01	0,7405	1,38	0,9556	1,75	0,9956
0,65	0,2079	1,02	0,7508	1,39	0,9580	1,76	0,9959
0,66	0,2236	1,03	0,7607	1,40	0,9603	1,77	0,9962
0,67	0,2395	1,04	0,7704	1,41	0,9624	1,78	0,9964
0,68	0,2557	1,05	0,7797	1,42	0,9645	1,79	0,9967
0,69	0,2721	1,06	0,7888	1,43	0,9665	1,80	0,9969
0,70	0,2887	1,07	0,7976	1,44	0,9683	1,81	0,9971

Scanned with OKEN Scanner

Tabelul 5, a

Valorile lui $F_{0.95}$ (testul F)

n_2	n_1 (numărul gradelor de libertate ale numărătorului)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18,5	19,0	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,0	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	17,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	2,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,32	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,23	2,20	2,18	2,15	2,13

Scanned with OKEN Scanner

24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,15	2,13	2,11
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	1,02	1,99	1,96	1,94
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99	1,96	1,93	1,91
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,05	2,09	2,04	2,00	1,97	1,94	1,94	1,89
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83	1,80	1,77	1,75
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82	1,79	1,76	1,73
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78	1,75	1,72	1,70
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,69
1000	3,85	3,00	2,61	2,28	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76	1,73	1,70	1,68

Scanned with OKEN Scanner

$F_{0,99}$

Tabelul 5, b

n_2	n_1 (numărul gradelor de libertate ale numărătorului)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0	26,9	26,9
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,4	14,3	14,3	14,2
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82	9,77	9,72
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	3,61	5,56	5,52
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,00	4,96
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,75	4,60	4,56
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25
12	9,33	6,93	5,95	5,41	5,07	4,82	4,54	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31
18	8,29	6,01	5,09	4,59	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	2,12	3,07	3,02	2,98
23	7,88	5,66	4,76	4,26	3,91	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89

Scanned with OKEN Scanner

25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90	2,86	2,82
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,09	2,93	2,87	2,82	2,78
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,92	2,84	2,79	2,75
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,77	2,81	2,77	2,73
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80	2,74	2,70	2,66
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76	2,70	2,66	2,62
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72	2,67	2,62	2,58
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69	2,64	2,59	2,55
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64	2,59	2,54	2,50
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62	2,56	2,52	2,47
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60	2,54	2,50	2,45
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,72	2,64	2,58	2,53	2,48	2,44
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70	2,63	2,56	2,51	2,46	2,42
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53	2,47	2,42	2,38
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27
90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33	2,29	2,24
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,26	2,22
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33	2,25	2,20	2,16
150	6,81	4,75	3,92	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31	2,25	2,20	2,16
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22	2,17	2,13
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24	2,19	2,14	2,10
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22	2,17	1,12	2,07
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20	2,15	2,10	2,06

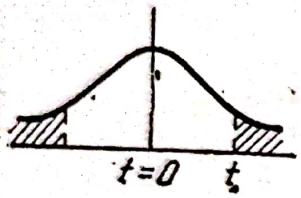
Scanned with OKEN Scanner

Valorilor lui $F_{0,975}$ (testul F)

Tabelul 5,c

n_2	n_1 (numărul gradelor de libertate ale numărătorului)														
	1	2	3	4	5	6	7	8	9	10	20	30	40	60	∞
1	648	800	864	900	922	937	948	957	963	969	993	1,001	1,006	1,010	1,018
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4	14,2	14,1	14,0	14,0	13,9
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,56	8,46	8,41	8,36	8,26
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,33	6,23	6,18	6,12	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,17	5,07	5,01	4,96	4,85
7	8,07	6,45	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,47	4,36	4,31	4,25	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,00	3,89	3,84	3,78	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,67	3,56	3,51	3,45	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,42	3,31	3,26	3,20	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,23	3,12	3,06	3,00	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,07	2,96	2,91	2,85	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	2,95	2,84	2,78	2,72	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,28	3,21	3,15	2,84	2,73	2,67	2,61	2,49
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,76	2,64	2,59	2,52	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,68	2,57	2,51	2,45	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,62	2,50	2,44	2,38	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,56	2,44	2,38	2,32	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,51	2,39	2,33	2,27	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,46	2,35	2,29	2,22	2,09
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,42	2,31	2,25	2,18	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,39	2,27	2,21	2,14	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,36	2,24	2,18	2,11	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,33	2,21	2,15	2,08	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,30	2,18	2,12	2,05	1,91
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,20	2,07	2,01	1,94	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,07	1,94	1,88	1,80	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	1,94	1,82	1,74	1,67	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	1,82	1,69	1,61	1,53	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,71	1,57	1,48	1,39	1,00

Scanned with OKEN Scanner



Tabelul 6

Valorile lui t (testul Student)

α reprezintă suma ariilor haşurate pentru valorile lui t date mai jos

s	$\alpha = 0{,}05$	$\alpha = 0{,}02$	$\alpha = 0{,}01$	$\alpha = 0{,}001$	s	$\alpha = 0{,}05$	$\alpha = 0{,}02$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
1	12,71	31,82	63,66	636,6	20	2,086	2,528	2,845	3,850
2	4,303	6,965	9,925	31,60	21	2,080	2,518	2,831	3,819
3	3,182	4,541	5,841	12,92	22	2,074	2,508	2,819	3,792
4	2,776	3,747	4,604	8,610	23	2,069	2,500	2,807	3,767
5	2,571	3,365	4,032	6,869	24	2,064	2,492	2,797	3,745
6	2,447	3,153	3,707	5,959	25	2,060	2,485	2,787	3,725
7	2,365	2,998	3,499	5,408	26	2,056	2,479	2,779	3,707
8	2,306	2,896	3,355	5,041	27	2,052	2,473	2,771	3,690
9	2,262	2,821	3,250	4,781	28	2,048	2,467	2,763	3,674
10	2,228	2,764	3,169	4,587	29	2,045	2,462	2,756	3,659
11	2,201	2,718	3,106	4,437	30	2,042	2,457	2,750	3,646
12	2,179	2,681	3,055	4,318	40	2,021	2,423	2,704	3,551
13	2,160	2,650	3,012	4,221	50	2,009	2,403	2,678	3,495
14	2,145	2,624	2,977	4,140	60	2,000	2,390	2,660	3,460
15	2,131	2,602	2,147	4,073	80	1,990	2,374	2,639	3,415
16	2,120	2,583	2,921	4,015	100	1,984	2,365	2,626	3,389
17	2,101	2,567	2,898	3,965	200	1,792	2,345	2,601	3,339
18	2,101	2,552	2,878	3,922	500	1,965	2,334	2,586	3,310
19	2,093	2,539	2,861	3,883	∞	1,960	2,326	2,576	3,291

Scanned with OKEN Scanner

Valorile lui Z, $Z = \frac{1}{2} \lg \frac{1+r}{1-r}$

r	z	r	z	r	z
0,00	0,000				
0,01	0,010	0,36	0,377	0,71	0,887
0,02	0,020	0,37	0,388	0,72	0,908
0,03	0,030	0,38	0,400	0,73	0,929
0,04	0,040	0,39	0,412	0,74	0,950
0,05	0,050	0,40	0,424	0,75	0,973
0,06	0,060	0,41	0,436	0,76	0,996
0,07	0,070	0,42	0,448	0,77	1,020
0,08	0,080	0,43	0,460	0,78	1,045
0,09	0,090	0,44	0,472	0,79	1,071
0,10	0,100	0,45	0,485	0,80	1,099
0,11	0,110	0,46	0,497	0,81	1,127
0,12	0,121	0,47	0,510	0,82	1,157
0,13	0,131	0,48	0,523	0,83	1,188
0,14	0,141	0,49	0,536	0,84	1,221
0,15	0,151	0,50	0,549	0,85	1,256
0,16	0,161	0,51	0,563	0,86	1,293
0,17	1,172	0,52	0,576	0,87	1,333
0,18	0,182	0,53	0,590	0,88	1,376
0,19	0,192	0,54	0,604	0,89	1,422
0,20	0,203	0,55	0,618	0,90	1,472
0,21	0,213	0,56	0,633	0,91	1,528
0,22	0,224	0,57	0,648	0,92	1,589
0,23	0,234	0,58	0,662	0,93	1,658
0,24	0,245	0,59	0,678	0,94	1,738
0,25	0,255	0,60	0,693	0,95	1,832
0,26	0,266	0,61	0,709	0,96	1,946
0,27	0,277	0,62	0,725	0,97	2,092
0,28	0,288	0,63	0,741	0,98	2,298
0,29	0,399	0,64	0,758	0,99	2,647
0,30	0,310	0,65	0,775		
0,31	0,321	0,66	0,793		
0,32	0,332	0,67	0,811		
0,33	0,343	0,68	0,829		
0,34	0,354	0,69	0,848		
0,35	0,365	0,70	0,867		

Scanned with OKEN Scanner

Tabelul 8

Valorile lui X_ε (testul lui van der Waerden)

$\varepsilon = 0{,}025$				$\varepsilon = 0{,}010$				$\varepsilon = 0{,}005$			
n	$n_1 - n_2$			n	$n_1 - n_2$			n	$n_1 - n_2$		
	0 sau 1	2 sau 3	4 sau 5		0 sau 1	2 sau 3	4 sau 5		0 sau 1	2 sau 3	4 sau 5
6	—	—	—	6	—	—	—	6	—	—	—
7	—	—	—	7	—	—	—	7	—	—	—
8	2,40	2,30	—	8	—	—	—	8	—	—	—
9	2,38	2,20	—	9	2,80	—	—	9	—	—	—
10	2,60	2,49	2,30	10	3,00	2,90	2,80	10	3,20	3,10	—
11	2,72	2,58	2,40	11	3,20	3,00	2,90	11	3,40	3,40	—
12	2,86	2,79	2,68	12	3,29	3,30	3,20	12	3,60	3,58	3,40
13	2,96	2,91	2,78	13	3,50	3,36	3,18	13	3,71	3,68	3,50
14	3,11	3,06	3,00	14	3,62	3,55	3,46	14	3,94	3,88	3,76
15	3,24	3,19	3,06	15	3,74	3,68	3,57	15	4,07	4,05	3,88
16	3,39	3,36	3,28	16	3,92	3,90	3,80	16	4,26	4,25	4,12
17	3,49	3,44	3,36	17	4,06	4,01	3,90	17	4,44	4,37	4,23
18	3,63	3,60	3,53	18	4,23	4,21	4,14	18	4,60	4,58	4,50
19	3,73	3,69	3,61	19	4,37	4,32	4,23	19	4,77	4,71	4,62
20	3,86	3,84	3,78	20	4,52	4,50	4,44	20	4,94	4,92	4,85
21	3,96	3,92	3,85	21	4,66	4,62	4,53	21	5,10	5,05	4,96
22	4,08	4,06	4,01	22	4,80	4,78	4,72	22	5,26	5,24	5,17
23	4,18	4,15	4,08	23	4,92	4,89	4,81	23	5,40	5,36	5,27
24	4,29	4,27	4,23	24	5,06	5,04	4,99	24	5,55	5,53	5,48
25	4,39	4,36	4,30	25	5,18	5,14	5,08	25	5,68	5,65	5,58

Л. Н Большев, Н. Б. Смирнов
Таблицы Математической статистики, издательство „Наука" Москва, 1965

Scanned with OKEN Scanner

Tabelul 8 (continuare)

n	$\varepsilon = 0{,}0250$			n	$\varepsilon = 0{,}010$			n	$\varepsilon = 0{,}005$		
	$n_1 - n_2$				$n_1 - n_2$				$n_1 - n_2$		
	0 sau 1	2 sau 3	4 sau 5		0 sau 1	2 sau 3	4 sau 5		0 sau 1	2 sau 3	4 sau 4
26	4,50	4,48	4,44	26	5,30	5,29	5,24	26	5,83	5,81	5,76
27	4,59	4,56	4,51	27	5,42	5,39	5,33	27	5,95	5,92	5,85
28	4,69	4,68	4,64	28	5,54	5,52	5,48	28	6,09	6,07	6,03
29	4,78	4,76	4,72	29	5,65	5,62	5,57	29	6,22	6,19	6,13
30	4,88	4,87	4,84	30	5,77	5,75	5,72	30	6,35	6,34	6,30
31	4,97	4,95	4,91	31	5,87	5,85	5,80	31	6,47	6,44	6,39
32	5,07	5,06	5,03	32	5,99	5,97	5,94	32	6,60	6,58	6,55
33	5,15	5,13	5,10	33	6,09	6,07	6,02	33	6,71	6,69	6,64
34	5,25	5,24	5,21	34	6,20	6,19	6,16	34	6,84	6,82	6,79
35	5,33	5,31	5,28	35	6,30	6,28	6,24	35	6,95	6,92	6,88
36	5,42	5,41	5,38	36	6,40	6,39	6,37	36	7,06	7,05	7,02
37	5,50	5,48	5,45	37	6,50	6,48	6,45	37	7,17	7,15	7,11
38	5,59	5,58	5,55	38	6,60	6,59	6,57	38	7,28	7,27	7,25
39	5,67	5,65	5,62	39	6,70	6,68	6,65	39	7,39	7,37	7,33
40	5,75	5,74	5,72	40	6,80	6,79	6,77	40	7,50	7,49	7,47
41	5,83	5,81	5,79	41	6,89	6,88	6,85	41	7,62	7,60	7,56
42	5,91	5,90	5,88	42	6,99	6,98	6,96	42	7,72	7,71	7,69
43	5,99	5,97	5,95	43	7,08	7,07	7,04	43	7,82	7,81	7,77
44	6,06	6,06	6,04	44	7,17	7,17	7,14	44	7,93	7,92	7,90
45	6,14	6,12	6,10	45	7,26	7,25	7,22	45	8,02	8,01	7,98
46	6,21	6,21	6,19	46	7,35	7,35	7,32	46	8,13	8,12	8,10
47	6,29	6,27	6,25	47	7,44	7,43	7,40	47	8,22	8,21	8,18
48	6,36	6,35	6,34	48	7,53	7,52	7,50	48	8,32	8,31	8,29
49	6,43	6,42	6,39	49	7,61	7,60	7,57	49	8,41	8,40	8,37
50	6,50	6,50	6,48	50	7,70	7,69	7,68	50	8,51	8,50	8,48

Scanned with OKEN Scanner

Testul lui Cochran. Valorile critice de 1 % pentru statistica $Q = \dfrac{\max\limits_{1 \leqslant i \leqslant k} S_i^2}{\sum\limits_{1}^{k} S_i^2}$ **determinată pentru** k **estimaţii independente ale dispersiilor, fiecare avînd** ν **grade de libertate**

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9999	0,9950	0,9794	0,9586	0,9373	0,9172	0,8988	0,8823	0,8674	0,8539	0,7949	0,7067	0,6062	0,5000
3	9933	9423	8831	8335	7933	7606	7335	7107	6912	6743	6059	5153	4230	3333
4	9676	8643	7814	7212	6761	6410	6129	5897	5702	5536	4884	4057	3251	2500
5	0,9279	0,7885	0,6957	0,6329	0,5875	0,5531	0,5259	0,5037	0,4854	0,4697	0,4094	0,3351	0,2644	0,2000
6	8828	7218	6258	5635	5195	4866	4608	4401	4229	4084	3529	2858	2229	1667
7	8376	6644	5685	5080	4659	4347	4105	3911	3751	3616	3105	2494	1929	1429
8	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704	0,3522	0,3373	0,3248	0,2779	0,2214	0,1700	0,1250
9	7544	5727	4810	4251	3870	3592	3378	3207	3067	2950	2514	1992	1521	1111
10	7175	5358	4469	3934	3572	3308	3106	2945	2813	2704	2297	1811	1376	1000
12	0,6528	0,4751	0,3919	0,3428	0,3099	0,2861	0,2680	0,2535	0,2419	0,2320	0,1961	0,1535	0,1157	0,0833
15	5747	4069	3317	2882	2593	2386	2228	2104	2002	1918	1612	1251	0934	0667
20	4799	3297	2654	2288	2048	1877	1748	1646	1567	1501	1248	0960	0709	0500
24	0,4247	0,2871	0,2295	0,1970	0,1759	0,1608	0,1495	0,1406	0,1338	0,1283	0,1060	0,0810	0,0595	0,0417
30	3632	2412	1913	1635	1454	1327	1232	1157	1100	1054	0867	0658	0480	0333
40	2940	1915	1508	1281	1135	1033	0957	0898	0853	0816	0668	0503	0363	0250
60	0,2151	0,1371	0,1069	0,0902	0,0796	0,0722	0,0668	0,0625	0,0594	0,0567	0,0461	0,0344	0,0245	0,0167
120	1225	0759	0585	0489	0429	0387	0357	0334	0316	0302	0242	0178	0125	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Tabelele 9,a şi 9,b sînt luate din Л. Н. БОЛЬШЕВ, Н. В. СМИРНОВ. ТАБЛИЦЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ, ИЗДАТЕЛЬСТВО „НАУКА", МОСКВА, 1965

Scanned with OKEN Scanner

Tabelul 9,b

Testul lui Cochran. Valorile critice de 5% pentru statistică $Q = \dfrac{\max\limits_{1 \leqslant \nu \leqslant k} S_i^2}{\sum\limits_{1}^{k} S_i^2}$ **determinată pentru** k **estimații independente ale dispersiilor, fiecare având** ν **grade de libertate**

k \ ν	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,7880	0,7341	0,6602	0,5813	0,5000
3	9669	8709	7977	7457	7071	6771	6530	6333	6167	6025	5466	4748	4031	3333
4	9065	7679	6841	6287	5895	5598	5365	5175	5017	4884	4366	3720	3093	2500
5	0,8412	0,6838	0,5981	0,5440	0,5063	0,4783	0,4564	0,4387	0,4241	0,4118	0,3645	0,3066	0,2513	0,2000
6	7808	6161	5321	4803	4447	4184	3980	3817	3682	3568	3135	2612	2119	1667
7	7271	5612	4800	4307	3974	3726	3535	3384	3259	3154	2756	2278	1833	1429
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0,1250
9	6385	4775	4027	3584	3286	3067	2901	2768	2659	2568	2226	1820	1446	1111
10	6020	4450	3733	3311	3029	2823	2666	2541	2439	2353	2032	1655	1308	1000
12	0,5410	0,3924	0,3264	0,2880	0,2624	0,2439	0,2299	0,2187	0,2098	0,2020	0,1737	0,1403	0,1100	0,0833
15	4709	3346	2758	2419	2195	2034	1911	1815	1736	1671	1429	1144	0889	0667
20	3894	2705	2205	1921	1735	1602	1501	1422	1357	1303	1108	0879	0675	0500
24	0,3434	0,2354	0,1907	0,1656	0,1493	0,1374	0,1286	0,1216	0,1160	0,1113	0,0942	0,0743	0,0567	0,0417
30	2929	1980	1593	1377	1237	1137	1061	1002	0958	0921	0771	0604	0457	0333
40	2370	1576	1259	1082	0968	0887	0827	0780	0745	0713	0595	0462	0347	0250
60	0,1737	0,1131	0,0895	0,0765	0,0682	0,0623	0,0583	0,0552	0,0520	0,0497	0,0411	0,0316	0,0234	0,0167
120	0998	0632	0495	0419	0371	0337	0312	0292	0279	0266	0218	0165	0120	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Scanned with OKEN Scanner

Coli tipo 14,25

Tiparul executat sub comanda
nr. 156 la
Întreprinderea Poligrafică,
„13 Decembrie 1918",
str. Grigore Alexandrescu nr. 89—97,
București,
Republica Socialistă România

Scanned with OKEN Scanner